

Chapter 8

Carrier Synchronization

Marvin K. Simon and Jon Hamkins

Traditionally, carrier synchronization (sync) techniques have been developed assuming that the modulation format and signal constellation characteristics are known a priori. By modulation format we mean that the modulation index is chosen so that either the carrier is fully suppressed or a residual carrier component remains. By constellation characteristics we refer to the shape of the constellation, e.g., a circle for M -ary phase-shift keying (M -PSK) or a square for quadrature amplitude modulation (QAM), and its size in terms of the number of signal points it contains. Aside from knowing the modulation index and signal constellation structure, it is also customary to have knowledge of the data rate and type (e.g., non-return to zero (NRZ) versus Manchester code) since the true optimum design of the loop depends on this information.

In autonomous radio operation, the most optimistic situation would be that the receiver contain a carrier synchronization structure that is capable of tracking the carrier phase independently of the above-mentioned considerations. Unfortunately, this is not completely possible since, for example, a squaring loop (or equivalently a binary phase-shift keying (BPSK) Costas loop) cannot track a quadrature phase-shift keying (QPSK) modulation and likewise a 4th power loop (or equivalently a QPSK Costas loop, sometimes referred to as an in-phase–quadrature (I-Q) loop) cannot properly track a BPSK signal.¹ Nevertheless, while in principle each carrier synchronization loop developed for a given modulation format, constellation, and data rate/type has certain unique characteristics, they do share a number of similarities, e.g., a common front-end demodulator

¹The inability of a QPSK Costas loop to properly track a BPSK signal will be treated later on in the chapter since this is an issue that has not been widely discussed in the literature.

structure, that allows one to consider designs that could be operational in the absence of complete a priori knowledge of all of these characteristics. For example, if the modulation is restricted to the M -PSK class, then it is possible to construct a universal structure that performs the carrier synchronization function for all values of M . This structure is derived by first determining the maximum a posteriori (MAP) estimate of carrier phase based on an observation of the received signal, namely, M -PSK plus additive white Gaussian noise (AWGN), and then using this to motivate a closed-loop carrier synchronization loop. Such a structure, referred to as the MAP estimation loop, has been previously proposed in the literature for cases where the modulation is known beforehand [1]. In fact, it can be shown (see Appendix 8-A for a derivation for BPSK modulation) that, by making an analogy between the closed-loop bandwidth and the noise bandwidth of an integrate-and-dump (I&D) filter of duration equal to the observation time for the open-loop MAP estimate, the closed loop approaches the Cramer–Rao lower bound on the variance of an unbiased estimate of the phase of a modulated carrier.

Still further, if the modulation is known to be other than suppressed carrier, i.e., a modulation index less than $\pi/2$ rad, then it is still possible to exploit the power in both the data and residual carrier components for carrier-tracking purposes provided one has knowledge of the modulation index itself. Such knowledge could be derived noncoherently, i.e., in the absence of carrier synchronization, from a suitable modulation index estimator (to be discussed elsewhere in the monograph). Loops of this type have been referred to in the literature as hybrid carrier tracking loops and like their suppressed-carrier counterparts are motivated by the same MAP considerations.

In what follows, we shall primarily restrict ourselves to the class of M -PSK modulations with known data format (pulse shape) that once again could be determined by a separate data format classifier operating noncoherently (to be discussed in another chapter of the monograph). It is also possible with minor modification, e.g., by replacing the matched filters in the I and Q arms of the loop with simple low-pass filters, to make the carrier synchronizer operation somewhat independent of the exact pulse shape but not without some attendant loss in performance. In deriving a generic carrier synchronization structure for this class of modulations, we shall consider a system with fixed modulation bandwidth which implicitly implies a fixed data *symbol* rate for all values of M . This is consistent with the same assumption made for various other classifiers in other chapters of the monograph.

Although the MAP estimation loops mentioned above are optimum in the sense of yielding the best tracking performance as measured by the variance of the loop phase error, their implementation typically involves nonlinearities that depend on other system parameters, such as signal-to-noise ratio (SNR). To circumvent this dependence, the most convenient form for use in the autonomous

radio application is the simplification based on low SNR approximations applied to the nonlinearities inherent in the MAP phase estimate. When this is done, the error signal in the loop for M -PSK is of the form $\sin M\phi$, where ϕ is the loop phase error, which from simple trigonometry can be written as $\sin M\phi = 2 \sin [(M/2)\phi] \cos [(M/2)\phi]$. Thus, it is seen that the error signal in the loop for M -PSK is formed from the *product* of the error signal $\sin [(M/2)\phi]$ and the lock detector signal $\cos [(M/2)\phi]$ in the loop for $M/2$ -PSK modulation. This simple relationship forms the basis for implementing the universal structure and will be discussed and demonstrated later on in the chapter. For a further treatment of this subject, the reader is referred to [2].

8.1 Suppressed versus Residual Carrier Synchronization

In the past, carrier synchronization loops typically have fallen into two categories: those that track a discrete carrier, e.g., the phase-locked loop (PLL), and those that track a fully suppressed carrier, e.g., the Costas loop. A fully suppressed carrier comes about when a digital modulation is impressed on a carrier with a modulation index equal to $\pi/2$ rad, whereas a discrete (residual) carrier component appears in the spectrum when the modulation index is less than $\pi/2$ rad. For example, consider a binary modulation phase modulated onto a carrier with modulation index β , which in mathematical form is described by

$$s(t) = \sqrt{2P_t} \sin(\omega_c t + \beta m(t) + \theta_c) \quad (8-1)$$

where P_t is the total available transmitter power, ω_c is the radian carrier frequency, θ_c is the unknown carrier phase to be tracked, and $m(t) = \sum_{n=-\infty}^{\infty} c_n \times p(t - nT)$ is the data modulation with $p(t)$ the pulse shape, $\{c_n\}$ the random binary data taking on values ± 1 with equal probability, and T the data (baud) interval ($\mathcal{R} = 1/T$ is the data rate). Since for NRZ data $p(t)$ is a unit rectangle and for Manchester code $p(t)$ is a unit square wave, then because of the purely digital (± 1) nature of $m(t)$, by applying simple trigonometry to Eq. (8-1) we get

$$\begin{aligned} s(t, \theta_c) &= \sqrt{2P_t} \cos \beta \sin(\omega_c t + \theta_c) + \sqrt{2P_t} \sin \beta m(t) \cos(\omega_c t + \theta_c) \\ &= \sqrt{2P_c} \sin(\omega_c t + \theta_c) + \sqrt{2P_d} m(t) \cos(\omega_c t + \theta_c) \end{aligned} \quad (8-2)$$

where $P_c = P_t \cos^2 \beta$ denotes the power in the carrier (unmodulated) component and $P_d = P_t \sin^2 \beta$ denotes the power in the data (modulated) component. Since the power spectral density (PSD) of an NRZ-formatted signal is of the form

$(\sin \pi fT/\pi fT)^2$, then direct modulation of the carrier with such a waveform would, for any modulation index $\beta < \pi/2$, result in a discrete carrier occurring at the point of maximum energy (i.e., $f = 0$) in the data modulation spectrum. This in turn makes it difficult to extract carrier synchronization from the discrete component with the loop most commonly used for such purposes, namely, a PLL. The loss due to the overlapping spectrum is $1/(1 + 2E_s)$, where E_s is the symbol energy [3]. Because of this, the National Aeronautics and Space Administration (NASA) proximity-link standard [4] stipulates that direct modulation of a carrier with BPSK having NRZ formatting is always used in a suppressed-carrier mode, i.e., with $\beta = \pi/2$. On the other hand, since a Manchester-coded signal has a PSD of the form $[\sin^2(\pi fT/2)/(\pi fT/2)]^2$ that has a null at zero frequency, then it is quite natural to allow for insertion of a discrete carrier there, and thus a modulation index $\beta < \pi/2$ is certainly reasonable from a carrier-tracking standpoint.

Since a Manchester-coded waveform is equivalent to the product of an NRZ waveform and a unit square wave at the data rate, one can view the form of the signal in Eq. (8-1) for this case as direct modulation of a carrier with an NRZ data waveform that has first been modulated onto a square-wave subcarrier at the data rate. With that in mind, one could, as is often done (at the expense of an increase in bandwidth of the signal), achieve a non-suppressed-carrier mode of operation with an NRZ signal by first modulating it onto a square-wave² subcarrier (not necessarily at the data rate) prior to direct modulation of the carrier. When this is done, the signal takes the form

$$s(t, \theta_c) = \sqrt{2P_t} \sin(\omega_c t + \beta m(t) Sq(\omega_{sc} t) + \theta_c) \quad (8-3)$$

where ω_{sc} denotes the radian subcarrier frequency. Most of the discussion of this chapter will deal with the absence of subcarriers and, thus, unless otherwise specified, when considering a residual carrier mode of operation, we shall implicitly assume the presence of Manchester coding, whereas for suppressed-carrier operation we shall allow for either NRZ or Manchester formats.

8.2 Hybrid Carrier Synchronization

Despite the fact that a data-modulated suppressed-carrier signal component also exists in Eq. (8-2), it is often neglected in deriving carrier synchronization. In other words, for the case where the total transmitted power is divided between a discrete (unmodulated) carrier and a data-modulated suppressed carrier, the

² Often, a sine-wave subcarrier is used with the same purpose of shifting the PSD of the baseband modulation away from the origin to allow insertion of a discrete carrier.

carrier synchronization function is most often accomplished based on the discrete carrier component alone, i.e., with a PLL. Thus, since the power split between discrete and data-modulated carriers results in a carrier power, P_c , that is less than the total transmitted power, P_t , the loop is operating with an SNR less than that which is potentially available if one were to employ both signal components in the carrier synchronization process.

Since a PLL is a closed-loop synchronization scheme motivated by MAP estimation of the phase of a discrete carrier and a Costas loop is a closed-loop synchronization scheme motivated by MAP estimation of the phase of a fully suppressed carrier, one might anticipate that for a signal of the form in Eq. (8-2) the optimum (in the MAP sense) closed-loop scheme would be a combination (hybrid) of the two loops [5]. Indeed such is the case, as is illustrated by the following mathematical development.

Let the signal of Eq. (8-2) received in AWGN be denoted by

$$r(t) = s(t, \theta_c) + n(t) \quad (8-4)$$

Then the likelihood function (conditioned on the unknown phase and data) for the k th interval is given by

$$p(r_k | \theta_c, c_k) = C \exp \left\{ -\frac{1}{N_0} \left[\int_{kT}^{(k+1)T} \left(r(t) - \sqrt{2P_d} c_k p(t - kT) \sin(\omega_c t + \theta_c) - \sqrt{2P_c} \cos(\omega_c t + \theta_c) \right)^2 dt \right] \right\} \quad (8-5)$$

where N_0 is the single-sided noise power spectral density in W/Hz and C is a constant of proportionality. Averaging over the data and ignoring terms that are not decision-dependent gives

$$p(r_k | \theta_c) \cong \exp \left\{ \frac{2\sqrt{2P_c}}{N_0} \int_{kT}^{(k+1)T} r(t) \cos(\omega_c t + \theta_c) dt \right\} \times \cosh \left\{ \frac{2\sqrt{2P_d}}{N_0} \int_{kT}^{(k+1)T} r(t) p(t - kT) \sin(\omega_c t + \theta_c) dt \right\} \quad (8-6)$$

The log-likelihood function for a sequence of K bits is then

$$\Lambda(\theta_c) \cong \ln \prod_{k=0}^{K-1} p(r_k | \theta_c) = \sum_{k=0}^{K-1} \frac{2\sqrt{2P_c}}{N_0} \int_{kT}^{(k+1)T} r(t) \cos(\omega_c t + \theta_c) dt$$

$$+ \sum_{k=0}^{K-1} \ln \cosh \left\{ \frac{2\sqrt{2P_d}}{N_0} \int_{kT}^{(k+1)T} r(t) p(t - kT) \sin(\omega_c t + \theta_c) dt \right\} \quad (8-7)$$

Thus, the MAP open-loop estimate of θ_c , namely, $\hat{\theta}_c$, based on the K -bit observation of $r(t)$ is that value of θ_c that maximizes $\Lambda(\theta_c)$.

To obtain a closed-loop synchronizer motivated by the MAP estimation approach, one differentiates the log-likelihood function with respect to θ_c and uses this to form the error signal (to be nulled when $\theta_c = \hat{\theta}_c$) in the closed-loop configuration. Thus, differentiating $\Lambda(\theta_c)$ of Eq. (8-7) with respect to θ_c gives

$$\frac{d\Lambda(\theta_c)}{d\theta_c} \cong - \sum_{k=0}^{K-1} \frac{2\sqrt{2P_c}}{N_0} \int_{kT}^{(k+1)T} r(t) \sin(\omega_c t + \theta_c) dt$$

$$+ \sum_{k=0}^{K-1} \left(\frac{2\sqrt{2P_d}}{N_0} \int_{kT}^{(k+1)T} r(t) p(t - kT) \cos(\omega_c t + \theta_c) dt \right)$$

$$\times \tanh \left\{ \frac{2\sqrt{2P_d}}{N_0} \int_{kT}^{(k+1)T} r(t) p(t - kT) \sin(\omega_c t + \theta_c) dt \right\} \quad (8-8)$$

The expression in Eq. (8-8) suggests the hybrid closed loop illustrated in Fig. 8-1. As is typical in actual implementations, the hyperbolic tangent nonlinearity is approximated by either its large argument variant, namely, a signum function (bipolar hard-limiter), or its small argument variant, a linear function. In the former case, we obtain the so-called polarity-type Costas loop, whereas in the latter case we obtain the conventional Costas loop. Also, implicit in Fig. 8-1 is knowledge of the data rate and symbol synchronization, both of which are necessary to implement the matched arm filters, which are of the I&D type. In the next section, we discuss alternative implementations of these arm filters using low-pass filters (LPFs) that are suboptimum but that provide additional robustness to the implementation in terms of the absence of perfect knowledge of the data rate and actual pulse shape and as such do not require symbol synchronization information. In this regard, Fig. 8-2 is the equivalent structure to Fig. 8-1, now using passive arm filters.

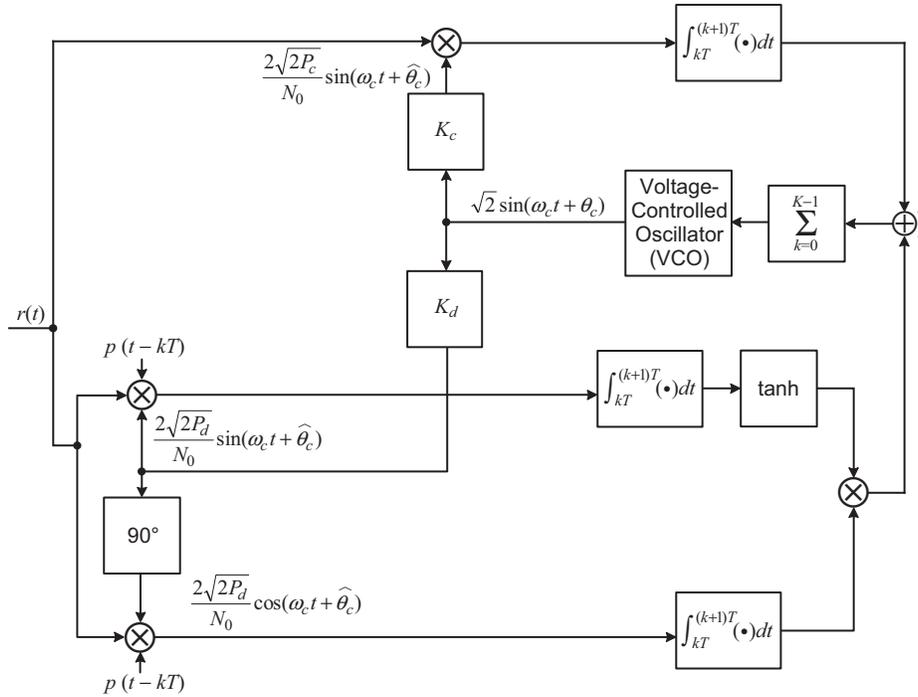


Fig. 8-1. Hybrid closed loop motivated by MAP estimation.

In theory, the gains K_c and K_d would be chosen proportional to $\sqrt{P_c}$ and $\sqrt{P_d}$, which in turn implies knowledge of the modulation index β . In the absence of such perfect knowledge, one would set the gains based on an estimate of β . Even in the absence of such information, one could possibly still use just the Costas-loop component of the hybrid loop alone since, under certain circumstances, it is capable of tracking a residual carrier signal whereas a data-modulated suppressed carrier cannot be tracked by a PLL. Rather than develop the conditions under which this is possible now, we delay this discussion until after we first compare the behavior of Costas loops using active (matched-filter) arm filters to those using passive low-pass filters. At that point, the behavior of the Costas loop with matched arm filters when tracking a residual carrier signal will simply become a special case of that discussion.

8.3 Active versus Passive Arm Filters

The most common measure of performance for a carrier synchronization loop is the variance of the phase error $\phi = \theta_c - \hat{\theta}_c$. For suppressed-carrier tracking loops such as the Costas loop (or the Costas-loop component of the hybrid loop),

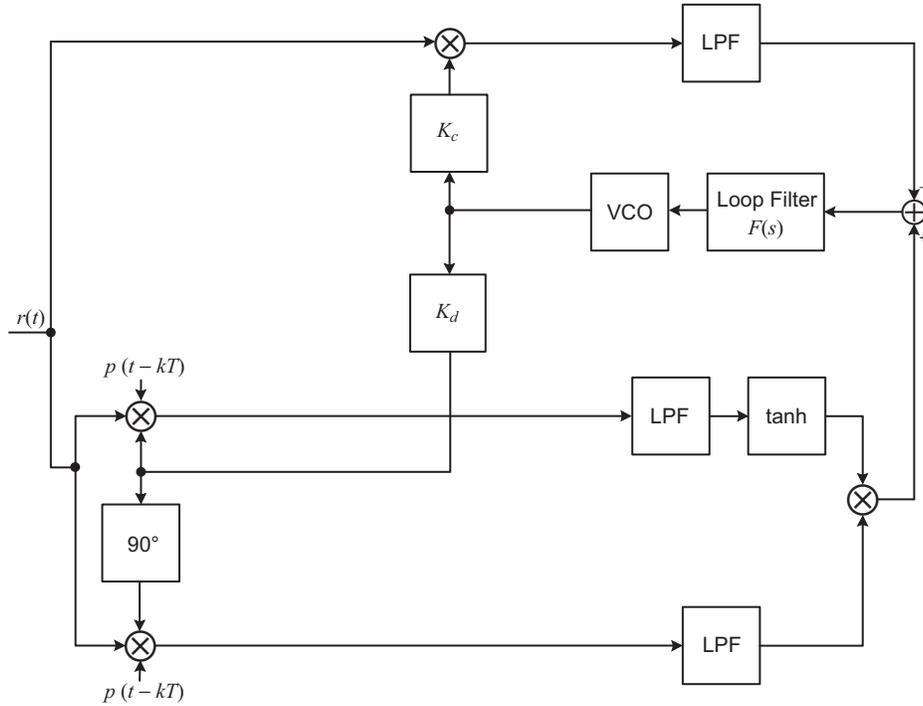


Fig. 8-2. MAP estimation loop for single channel, discrete carrier passive arm filter realization.

in the linear tracking region, the mean-squared phase error can be related to the loop SNR by

$$\sigma_{\phi}^2 = \frac{1}{\rho S_L} \quad (8-9)$$

where

$$\rho = \frac{P_d}{N_0 B_L} \quad (8-10)$$

with B_L denoting the single-sided loop bandwidth and S_L the so-called “squaring loss,” which reflects the additional penalty relative to the PLL loop SNR due to the squaring operation and is caused by the combination of signal \times signal ($S \times S$), signal \times noise ($S \times N$), and noise \times noise ($N \times N$) distortions. The exact nature of the squaring loss depends heavily on the nonlinearity implemented in the in-phase arm (i.e., hyperbolic tangent function or its small and large argument approximations) and the type (active versus passive) of arm filters in both the I and Q arms.

From the standpoint of performance, the optimum behavior is obtained using a hyperbolic tangent nonlinearity and matched (to the pulse shape) arm filters. For this case, the squaring loss is given by [6]

$$S_L = \frac{\left(\overline{\tanh [2R_d - \sqrt{2R_d X}]^X} \right)^2}{\overline{\tanh^2 [2R_d - \sqrt{2R_d X}]^X}}$$

$$= \exp(-R_d) \overline{\tanh(\sqrt{2R_d X}) \sinh(\sqrt{2R_d X})^X} \quad (8-11)$$

where X is a (0,1) Gaussian random variable (RV), the overbar indicates statistical averaging with respect to the Gaussian probability distribution of X , and $R_d \triangleq P_d T / N_0$ is the data SNR. For the polarity-type Costas loop ($\tanh x \cong \text{sgn } x$), Eq. (8-11) simplifies to

$$S_L = \text{erf}^2(\sqrt{R_d}) \quad (8-12)$$

where $\text{erf } x = (2/\sqrt{\pi}) \int_0^x \exp(-y^2) dy$ is the error function, whereas for the conventional Costas loop ($\tanh x \cong x$), we obtain

$$S_L = \frac{2R_d}{1 + 2R_d} \quad (8-13)$$

As a compromise between Eqs. (8-12) and (8-13), the hyperbolic tangent nonlinearity is often approximated by a saturated amplifier, i.e.,

$$\tanh x \cong \begin{cases} x, & |x| \leq 1 \\ \text{sgn } x, & |x| > 1 \end{cases} \quad (8-14)$$

whose squaring loss can also be obtained in closed form as

$$S_L = \frac{\left\{ \sqrt{\frac{R_d}{\pi}} [\exp(-A_1^2) - \exp(-A_2^2)] + \left(R_d + \frac{1}{2}\right) \text{erf } A_1 + \left(R_d - \frac{1}{2}\right) \text{erf } A_2 \right\}^2}{1 - \sqrt{\frac{R_d}{\pi}} \left\{ (1 - 2R_d) \exp(-A_1^2) + (1 + 2R_d) \exp(-A_2^2) \right\} + \left[R_d(1 + 2R_d) - \frac{1}{2} \right] [\text{erf } A_1 + \text{erf } A_2]}$$

$$A_1 \triangleq \frac{1 + 2R_d}{2\sqrt{R_d}}, \quad A_2 \triangleq \frac{1 - 2R_d}{2\sqrt{R_d}} \quad (8-15)$$

Figure 8-3 is a plot of the squaring losses in Eqs. (8-11), (8-12), (8-13), and (8-15) versus R_d in dB. We observe that, depending on the value of R_d , the polarity-type and conventional Costas loops trade performance in terms of which is superior, whereas for all values of R_d , the loop implemented with the hyperbolic tangent nonlinearity provides the best performance (minimum squaring loss) with the performance of the saturated amplifier nonlinearity virtually identical to it. Once again we remind the reader that the performances predicted by Eqs. (8-11), (8-12), (8-13), and (8-15) require the implementation of matched arm filters, which in turn require knowledge of the data rate/type and also symbol synchronization.

Before leaving the discussion of Costas loops with active arm filters, it is of interest to compare the performance (phase error variance) of the loop with the Cramer–Rao bound [7] on the variance of an unbiased estimator of the phase of a modulated BPSK carrier. The derivation of such a bound is given in Appendix 8-A, where it is shown that

$$\sigma_\phi^2 \geq \begin{cases} \frac{1}{K(2R_d)^2}, & R_d \text{ small} \\ \frac{1}{K(2R_d)}, & R_d \text{ large} \end{cases} \quad (8-16)$$

with, consistent with the notation used earlier in the chapter, K the number of bits in the observation. For the I-Q Costas loop at low SNR, we can rewrite Eq. (8-9) combined with Eqs. (8-10) and (8-13) as

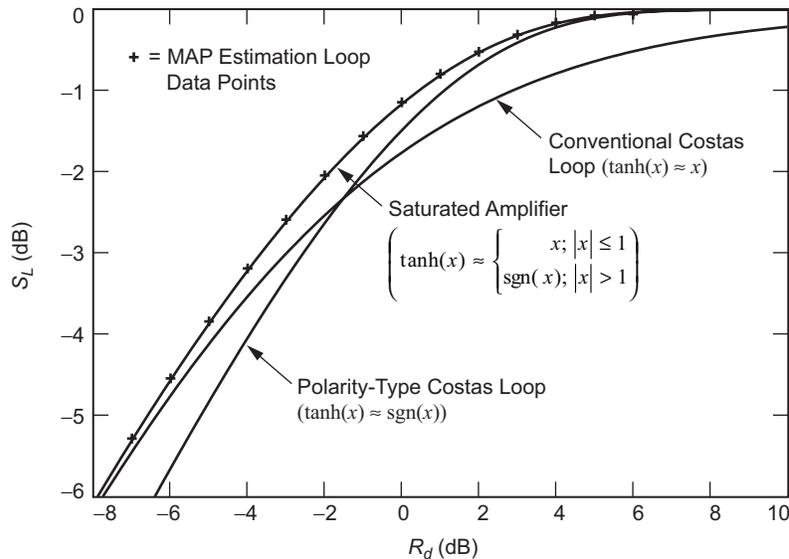


Fig. 8-3. A comparison of the squaring-loss performance of the MAP estimation loop with several practical implementations; BPSK.

$$\sigma_\phi^2 = \frac{N_0 B_L (1 + 2R_d)}{P_d (2R_d)} = \frac{2B_L T (1 + 2R_d)}{(2R_d)^2} \cong \frac{B_L (2KT)}{K (2R_d)^2} \quad (8-17)$$

Similarly, for high SNR we can rewrite Eq. (8-9) combined with Eqs. (8-10) and (8-12) as

$$\sigma_\phi^2 = \frac{N_0 B_L}{P_d \operatorname{erf}^2(\sqrt{R_d})} = \frac{2B_L T}{2R_d \operatorname{erf}^2(\sqrt{R_d})} \cong \frac{B_L (2KT)}{K (2R_d)} \quad (8-18)$$

Comparing Eqs. (8-17) and (8-18) with the Cramer–Rao bounds in Eq. (8-16), we observe that the performance of the I-Q Costas loop approaches these bounds at low and high SNR if we make the equivalence between the loop bandwidth and the noise bandwidth of an I&D of duration equal to the observation time, i.e., set $B_L = 1/2KT$. Thus, in conclusion, while for a fixed bandwidth and data rate the I&D Costas loop asymptotically behaves inverse linearly with SNR at high SNR, it has an asymptotic inverse square-law behavior with SNR at low SNR. In both cases, however, the behavior is inverse linear with the duration of the observation.

As intimated previously, it is possible to implement the arm filters of the Costas loop in passive form, thereby eliminating the need for symbol synchronization prior to obtaining carrier synchronization. Furthermore, as we shall see momentarily, in the absence of exact data rate information, the passive arm filter implementation is robust in that its performance is quite insensitive to a large variation of the data rate in one direction relative to the optimum choice of arm filter bandwidth. It is also possible to design the arm filters with different noise bandwidths ([8] suggests removing the quadrature arm filter completely), which has the advantage of improving the acquisition capability of the loop and also reducing its tendency to false lock but is accompanied by a penalty in tracking performance (as measured by squaring loss). For high-detection SNR, the additional squaring-loss penalty is quite small, and thus this technique could result in a significant overall performance advantage. For the current discussion, we shall assume that the two arm filters have identical designs.

Consider the Costas loop with LPFs having transfer function $G(s)$ (s is the Laplace transform operator) illustrated in Fig. 8-4. When operating in the linear tracking region, the mean-squared error is given by Eq. (8-9), where the squaring loss is now given by [9]

$$S_L = \frac{K_2^2}{K_4 + K_L \frac{B_i/\mathcal{R}}{2R_d}} \quad (8-19)$$

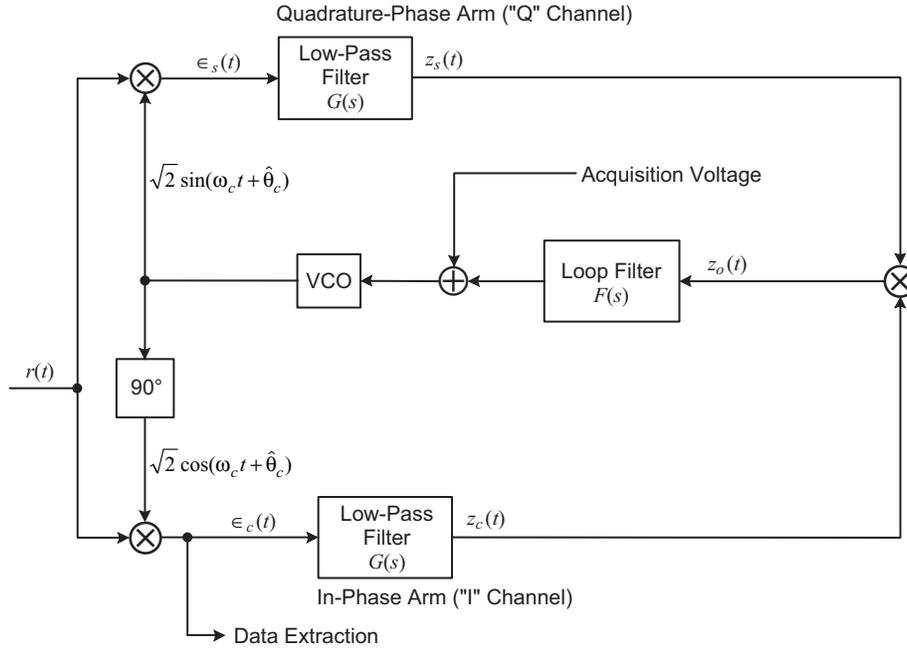


Fig. 8-4. The Costas loop with passive arm filters.

where

$$K_L = \frac{\int_{-\infty}^{\infty} |G(2\pi f)|^4 df}{\int_{-\infty}^{\infty} |G(2\pi f)|^2 df} \quad (8-20)$$

is a constant that depends only on the arm filter type,

$$K_l = \int_{-\infty}^{\infty} S_m(f) |G(2\pi f)|^l df, \quad l = 2, 4 \quad (8-21)$$

with

$$S_m(f) = \frac{1}{T} |P(j2\pi f)|^2 \quad (8-22)$$

the power spectral density of the modulation [$P(j2\pi f)$ is the Fourier transform of the pulse shape $p(t)$] and

$$B_i = \int_{-\infty}^{\infty} |G(j2\pi f)|^2 df \quad (8-23)$$

the two-sided noise bandwidth of the arm filters.³ Typical values of the quantities in Eqs. (8-20) through (8-23) for the class of N -pole Butterworth filters and several data formats are given below:

Filter Transfer Function:

$$|G(2\pi f)|^2 = \frac{1}{1 + \left(\frac{f}{f_{3 \text{ dB}}}\right)^{2N}} \quad (8-24)$$

$$B_i = \left(\frac{N}{\pi} \sin \frac{\pi}{2N}\right)^{-1} f_{3 \text{ dB}}$$

Data Modulation Power Spectral Density:

$$\text{NRZ : } S_m(f) = T \frac{\sin^2 \pi f T}{(\pi f T)^2} \quad (8-25)$$

$$\text{Manchester : } S_m(f) = T \frac{\sin^4 \pi f T / 2}{(\pi f T / 2)^2} \quad (8-26)$$

$$\text{Square Wave : } S_m(f) = \frac{1}{4} \left(\frac{4}{\pi}\right)^2 \sum_{k=-\infty}^{\infty} \frac{1}{(2k-1)^2} \delta\left(f - \frac{2k-1}{2T}\right) \quad (8-27)$$

The square wave in Eq. (8-27) has period $2T$.

Evaluation of K_L :

$$K_L = \frac{2N-1}{2N} \quad (8-28)$$

Evaluation of K_2, K_4 :

(a) Single-Pole Butterworth Filter ($N = 1$), NRZ Data:

³We assume that the arm filter transfer function is normalized such that $G(0) = 1$.

$$K_2 = 1 - \frac{1}{2B_i/\mathcal{R}} [1 - \exp(-2B_i/\mathcal{R})]$$

$$K_4 = 1 - \frac{3 - (3 + 2B_i/\mathcal{R}) \exp(-2B_i/\mathcal{R})}{4B_i/\mathcal{R}}$$
(8-29)

(b) Two-Pole Butterworth Filter ($N = 2$), NRZ Data:

$$K_2 = 1 - \frac{1}{4B_i/\mathcal{R}} \left\{ 1 - \exp\left(\frac{-2B_i}{\mathcal{R}}\right) \left[\cos\left(\frac{2B_i}{\mathcal{R}}\right) - \sin\left(\frac{2B_i}{\mathcal{R}}\right) \right] \right\}$$

$$K_4 = 1 - \frac{5 - \left\{ 4\left(\frac{B_i}{\mathcal{R}}\right) \cos\left(\frac{2B_i}{\mathcal{R}}\right) + 5 \left[\cos\left(\frac{2B_i}{\mathcal{R}}\right) - \sin\left(\frac{2B_i}{\mathcal{R}}\right) \right] \right\} \exp\left(\frac{-2B_i}{\mathcal{R}}\right)}{16B_i/\mathcal{R}}$$

(c) Single-Pole Butterworth Filter ($N = 1$), Manchester Data:

$$K_2 = 1 - \frac{1}{2B_i/\mathcal{R}} [3 - 4 \exp(-B_i/\mathcal{R}) + \exp(-2B_i/\mathcal{R})]$$

$$K_4 = 1 - \frac{9 - 4(3 + B_i/\mathcal{R}) \exp(-B_i/\mathcal{R}) + (3 + 2B_i/\mathcal{R}) \exp(-2B_i/\mathcal{R})}{4B_i/\mathcal{R}}$$
(8-30)

(d) Two-Pole Butterworth Filter ($N = 2$), Manchester Data:

$$K_2 = 1 - \frac{1}{4B_i/\mathcal{R}} \left\{ 3 - 4 \exp\left(\frac{-B_i}{\mathcal{R}}\right) \left[\cos\left(\frac{B_i}{\mathcal{R}}\right) - \sin\left(\frac{B_i}{\mathcal{R}}\right) \right] \right.$$

$$\left. + \exp\left(\frac{-2B_i}{\mathcal{R}}\right) \left[\cos\left(\frac{2B_i}{\mathcal{R}}\right) - \sin\left(\frac{2B_i}{\mathcal{R}}\right) \right] \right\}$$

$$K_4 = 1 - \frac{15 - \left\{ 8\left(\frac{B_i}{\mathcal{R}}\right) \cos\left(\frac{B_i}{\mathcal{R}}\right) + 20 \left[\cos\left(\frac{B_i}{\mathcal{R}}\right) - \sin\left(\frac{B_i}{\mathcal{R}}\right) \right] \right\} \exp\left(\frac{-B_i}{\mathcal{R}}\right)}{16B_i/\mathcal{R}}$$

$$- \frac{\left\{ 4\left(\frac{B_i}{\mathcal{R}}\right) \cos\left(\frac{2B_i}{\mathcal{R}}\right) + 5 \left[\cos\left(\frac{2B_i}{\mathcal{R}}\right) - \sin\left(\frac{2B_i}{\mathcal{R}}\right) \right] \right\} \exp\left(\frac{-2B_i}{\mathcal{R}}\right)}{16B_i/\mathcal{R}}$$
(8-31)

(e) Single-Pole Butterworth Filter ($N = 1$), Square Wave:

$$K_2 = 1 - \frac{1}{B_i/\mathcal{R}} \tanh\left(\frac{B_i}{\mathcal{R}}\right)$$

$$K_4 = 1 + \frac{1}{2} \operatorname{sech}^2\left(\frac{B_i}{\mathcal{R}}\right) - \frac{3}{2B_i/\mathcal{R}} \tanh\left(\frac{B_i}{\mathcal{R}}\right)$$

Finally, using a partial fraction expansion technique, closed-form expressions for the squaring loss, specifically, the parameters K_L, K_2, K_4 , were derived [10] for a general class of passive arm filters whose transfer function is characterized by simple, but in general complex, poles.

The numerator of Eq. (8-19) reflects the $S \times S$ distortion whereas the two terms of the denominator reflect the $S \times N$ and $N \times N$ distortions, respectively. As the arm filter bandwidth narrows, the decrease in the $S \times S$ term dominates, whereas when the arm filter widens, the increase in the $S \times N$ term dominates. Thus, for a fixed filter type and data modulation format, K_L, K_2 , and K_4 are only a function of the ratio of arm filter bandwidth to data rate B_i/\mathcal{R} , and thus a plot of S_L in dB versus B_i/\mathcal{R} will reveal an optimum value in the sense of maximum S_L .⁴ As an example, Figs. 8-5 and 8-6 illustrate such plots for one- and two-pole Butterworth arm filters and Manchester-coded data. We observe that, over a large range of detection SNRs, the squaring loss is extremely sensitive to values of B_i/\mathcal{R} less than the optimum value, whereas it is rather insensitive to values of B_i/\mathcal{R} greater than the optimum value. Thus, in the absence of exact information about the data rate, one can design the loop arm filters so that the optimum B_i/\mathcal{R} ratio corresponds to the *maximum* expected data rate whereupon operation at data rates considerably less than the maximum would result in only a small squaring-loss penalty. We remind the reader that, with active arm filters of the I&D type, one does not have this design flexibility since the arm filter bandwidth of such a filter is fixed at $B_i = 1/T = \mathcal{R}$, and thus the squaring loss is also fixed at its value given by Eq. (8-13). Figure 8-7 is a comparison of the squaring-loss behavior for one- and two-pole Butterworth filters at two different detection SNRs. While the two-pole filter has a slightly better optimum squaring-loss performance, it is also more sensitive to data rate variation above the optimum value than is the one-pole filter. Thus, we see that the design of the arm filter is a trade-off between the optimum performance and sensitivity to data rate variation.

The next question is: How much do we sacrifice in performance by using a passive arm filter rather than the active one (matched filter)? Figure 8-8 illustrates a comparison between the optimum squaring-loss performances of the

⁴ Note that since $S_L \leq 1$, maximizing S_L is equivalent to minimizing the squaring loss in dB.

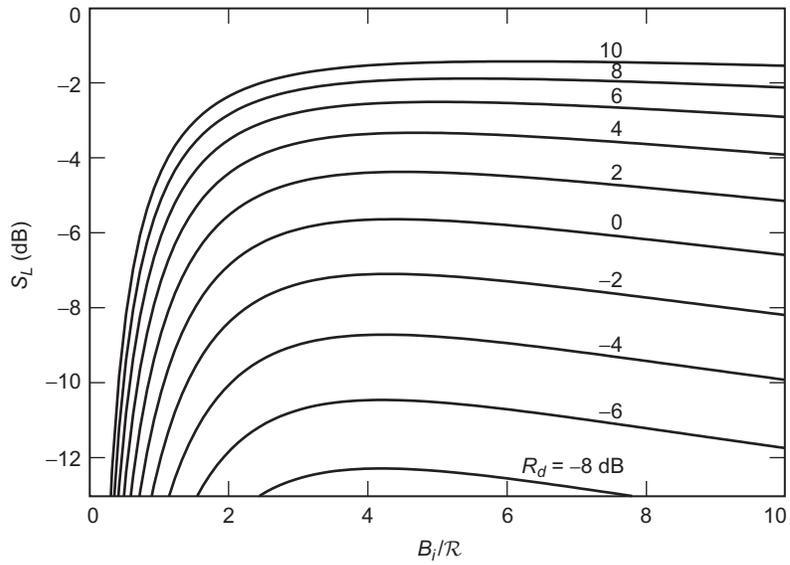


Fig. 8-5. Squaring-loss performance of Costas loop with single-pole Butterworth arm filters; Manchester-coded data.

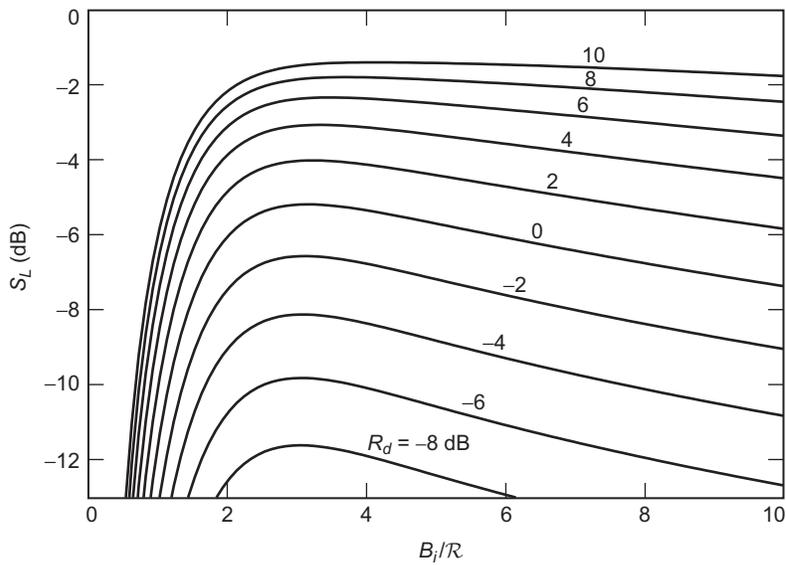


Fig. 8-6. Squaring-loss performance of Costas loop with two-pole Butterworth arm filters; Manchester-coded data.

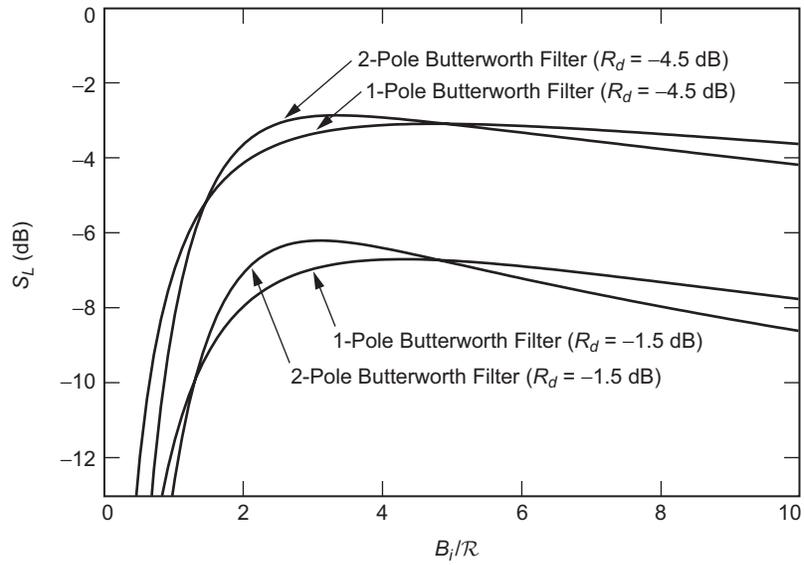


Fig. 8-7. A comparison of the squaring-loss performance of Costas loops with single- and two-pole arm filters at two different SNRs.

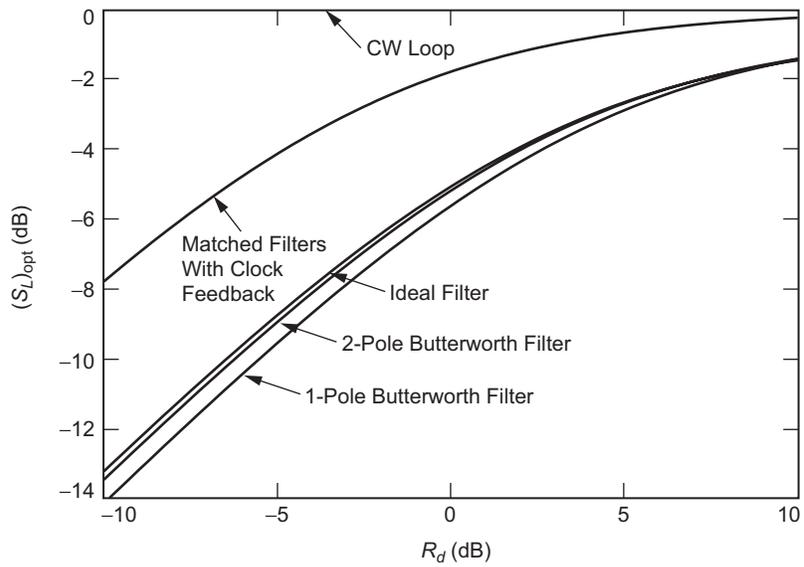


Fig. 8-8. A comparison of the squaring-loss performance of Costas loops with active and passive arm filters.

matched filter with perfect symbol synchronization, one- and two-pole Butterworth filters, and an ideal (brick wall) filter which corresponds to a Butterworth filter with $N \rightarrow \infty$. The curves are plotted for a Manchester data format. At high SNR, the optimum squaring-loss performance of the passive filters becomes independent of the number of filter poles, and at a detection SNR of 10 dB it is about 1 dB worse than that of the perfectly symbol-synchronized matched filter. We also note in passing that the performance of the matched-filter implementation is independent of the data format and is given by Eq. (8-13).

We now return to a point made earlier, namely, the ability of a Costas loop to successfully track a residual carrier BPSK signal, including the extreme case of a totally unmodulated carrier. When both a data-modulated and an unmodulated carrier component are simultaneously present at the input to a Costas loop, the two components tend to oppose each other at the error signal point in the loop. In fact, based on an analysis of a similar situation [11], it can be shown that there exists a critical modulation index at which the signal component of the loop error signal (loop S-curve) degenerates to zero, in which case the loop will not track at all. This critical modulation index is given by

$$\beta^* = \cot^{-1} \sqrt{K_2} \quad (8-32)$$

where K_2 is defined for a particular modulation format and arm filter type in Eq. (8-21). For a modulation index *greater* than β^* , the loop S-curve has the usual $\sin 2\phi$ characteristic and has stable lock points at $\phi = \pm n\pi$, $n = 0, 1, 2, \dots$, which corresponds to the desired operation of the loop (assuming that one takes measures to resolve the normal 180-deg phase ambiguity, such as differential encoding of the input data). On the other hand, for a modulation index *less* than β^* (which of course includes the limiting case of $\beta = 0$, a totally unmodulated carrier), the loop S-curve has a $-\sin 2\phi$ characteristic and has stable lock points at $\phi = \pm(2n+1)\pi/2$, $n = 0, 1, 2, \dots$ that, from a data detection standpoint, is an undesirable operating condition. Of course, if one knew the modulation index was in the region $\beta < \beta^*$, one could always insert a -1 gain in either the I or Q arm of the loop just prior to the I-Q multiplier, which would thereby invert the loop S-curve and reestablish the lock points at their desired location.

To quantify the degradation in performance in the presence of a residual carrier, it is straightforward to modify the results in [10] to show that under such conditions the squaring loss is given by

$$S_L = \frac{(K_2 \sin^2 \beta - \cos^2 \beta)^2}{\cos^2 \beta + K_4 \sin^2 \beta + \frac{1}{2} R_t (\sin^2 2\beta) \left(\frac{S_m(0)}{T} \right) + K_L \frac{B_i/R}{2R_t}} \quad (8-33)$$

where $R_t \triangleq P_t T / N_0$ is the total SNR. Since for Manchester coding $S_m(0) = 0$, Eq. (8-33) simplifies to

$$S_L = \frac{(K_2 \sin^2 \beta - \cos^2 \beta)^2}{\cos^2 \beta + K_4 \sin^2 \beta + K_L \frac{B_i/R}{2R_t}} \quad (8-34)$$

Next, for the discrete (unmodulated) carrier case wherein $\beta = 0$ ($P_t = P_c$), Eq. (8-33) becomes

$$S_L = \frac{1}{1 + K_L \frac{N_0 B_i}{2P_c}} \quad (8-35)$$

Note that, unlike the data-modulated case, there is no optimum arm filter bandwidth-to-data rate ratio since S_L is independent of the data rate. Furthermore, the squaring loss is smallest (S_L is maximum) for an unmodulated carrier. However, since, for any finite arm filter bandwidth, $S_L < 1$, then from Eq. (8-9), the phase error variance will always exceed that which arises from a PLL as alluded to previously.

When the data modulation is an NRZ-modulated square-wave subcarrier and as such the transmitted signal is characterized by Eq. (8-3), then the behavior of the Costas loop in the presence of a residual carrier is somewhat different from that described above. To illustrate this difference in the simplest way, we assume perfect subcarrier demodulation and once again active (I&D) arm filters. Under these circumstances, the I and Q I&D outputs would be given by

$$\begin{aligned} z_{ck} &= \int_{kT}^{(k+1)T} r(t) \left[\sqrt{2} \cos(\omega_c t + \hat{\theta}_c) S q(\omega_{sc} t) \right] dt \\ &= \sqrt{P_c} \sin \phi \int_{kT}^{(k+1)T} S q(\omega_{sc} t) dt + \sqrt{P_d} c_k \cos \phi \int_{kT}^{(k+1)T} S q^2(\omega_{sc} t) dt + N_c \\ &= \sqrt{P_d T} c_k \cos \phi + N_c \end{aligned} \quad (8-36)$$

$$\begin{aligned} z_{sk} &= \int_{kT}^{(k+1)T} r(t) \left[-\sqrt{2} \sin(\omega_c t + \hat{\theta}_c) S q(\omega_{sc} t) \right] dt \\ &= -\sqrt{P_c} \cos \phi \int_{kT}^{(k+1)T} S q(\omega_{sc} t) dt + \sqrt{P_d} c_k \sin \phi \int_{kT}^{(k+1)T} S q^2(\omega_{sc} t) dt + N_s \\ &= \sqrt{P_d T} c_k \sin \phi + N_s \end{aligned}$$

where N_c, N_s are again independent zero-mean Gaussian RVs with variance $N_0T/2$ and we have assumed a unit square-wave subcarrier with either an integer number of subcarrier cycles per bit or instead a large ratio of subcarrier frequency to bit rate. We observe that aside from the reduction of the power in the data signal from the total power P_t to $P_d = P_t \sin^2 \beta$, the I&D outputs in Eq. (8-36) are *identical* to what would be obtained for the same Costas loop operating in the conventional suppressed-carrier mode. Thus, we conclude (at least under the ideal assumptions made) that, in the case of an NRZ-modulated square-wave subcarrier, *the presence of a discrete carrier component does not degrade the performance of the loop other than to reduce the effective power in the data component by $\sin^2 \beta$* . Before leaving this subject, we also mention that since, as previously mentioned in Section 8-1, a Manchester-coded BPSK can be viewed as an NRZ data waveform modulated onto a square-wave subcarrier at the data rate prior to modulation onto the carrier, then, since the above arguments are independent of the subcarrier frequency, the same conclusion would also be true for this case when active (matched) filters are used in the I and Q arms of the Costas loop.

Finally, it is natural to ask whether there is an optimum passive arm filter type in the sense of minimizing the mean-squared phase error. In particular, one seeks a solution for $|G(2\pi f)|^2$ that minimizes

$$S_L^{-1} = \frac{P_d \int_{-\infty}^{\infty} S_m(f) |G(2\pi f)|^4 df + \frac{N_0}{2} \int_{-\infty}^{\infty} |G(2\pi f)|^2 df}{\left\{ \int_{-\infty}^{\infty} S_m(f) |G(2\pi f)|^2 df \right\}^2} \quad (8-37)$$

Using the method of Lagrange multipliers, it can be shown that the optimum arm filter magnitude-squared transfer function (which may or may not be physically realizable) is given by

$$|G_{opt}(2\pi f)|^2 = \lambda \frac{S_m(f)}{S_m(f) + \frac{N_0}{P_d}} \quad (8-38)$$

where λ is a Lagrange multiplier. Note that for low SNR ($P_d/N_0 \ll 1$), we obtain

$$G_{opt}(2\pi f) = \sqrt{\lambda \frac{P_d}{N_0}} [S_m(f)]^+ \quad (8-39)$$

where the “+” superscript refers to the part of $S_m(f)$ with poles in the left half-plane and as such represents the matched-filter solution.

8.4 Carrier Synchronization of Arbitrary Modulations

8.4.1 MPSK

In this section, we begin by continuing to consider the class of M -PSK modulations, where the value of $M = 2^m$ is now arbitrary. In an earlier section, we considered a carrier synchronization closed loop for BPSK that was motivated by the MAP estimation technique. Such an “optimum” loop was in the form of an I-Q structure with a hyperbolic tangent nonlinearity in its in-phase arm that resulted from using the gradient of the likelihood function as an error-control signal. Applying the MAP estimation technique for values of $M > 2$ [1,2], it can be shown that the derivative of the log-likelihood ratio has the form

$$\frac{d\Lambda(\theta_c)}{d\theta_c} \cong \sum_{k=0}^{K-1} \sum_{l=0}^{m-2} \frac{C_l V \tanh(C_l U) - S_l U \tanh(S_l V)}{1 + \sum_{\substack{n=0 \\ n \neq l}}^{m-2} \frac{\cosh(C_n U) \cosh(S_n V)}{\cosh(C_l U) \cosh(S_l V)}} \quad (8-40)$$

where

$$C_l \triangleq \cos \frac{(2l+1)\pi}{M}$$

$$S_l \triangleq \sin \frac{(2l+1)\pi}{M} \quad (8-41)$$

and

$$U = \frac{2\sqrt{2P_d}}{N_0} \int_{kT}^{(k+1)T} r(t) p(t - kT) \sin(\omega_c t + \theta_c) dt$$

$$V = \frac{2\sqrt{2P_d}}{N_0} \int_{kT}^{(k+1)T} r(t) p(t - kT) \cos(\omega_c t + \theta_c) dt \quad (8-42)$$

Once again using the derivative of the log-likelihood function to motivate an error signal in a closed-loop implementation, the carrier synchronization loop that results is again an I-Q structure; however, the nonlinearity no longer resides only in the in-phase arm and, furthermore, as can be seen from Eq. (8-40), is consider-

ably more complicated than a simple hyperbolic tangent function.⁵ An example of such a closed loop is illustrated in Fig. 8-9 for 8-PSK. We also have seen previously that, by approximating the nonlinearity for small and large arguments, one arrives at structures that are synonymous with well-known synchronization schemes for BPSK and approach the performance of the true MAP-motivated loop at low and high SNRs. Indeed, for $M > 2$, one can take the same approach and arrive at universal structures that lend themselves to simple implementation.

For large arguments we again use the approximation $\tanh x \cong \operatorname{sgn} x$ and in addition

$$\cosh x \cong \frac{1}{2} \exp(|x|) \quad (8-43)$$

Although not immediately obvious, these approximations lead to a closed-loop synchronizer that incorporates the optimum M -PSK symbol detector (MAP phase estimator) in its structure. An example of this is illustrated in Fig. 8-10 for 8-PSK and thus corresponds to the high SNR approximation of Fig. 8-9. For QPSK, the structure is somewhat simpler, involving hard-limiters (signum functions) in each of the I and Q arms as well as a crossover processing between the inputs and outputs of these nonlinearities to form the error signal [5] (see Fig. 8-11 for the passive arm filter implementation).

For small arguments, if one tries to use only the first term in the power series expansion of $\tanh x$, it can be shown [5,6] that for $M > 2$ the derivative of the log-likelihood function in Eq. (8-40) becomes equal to zero for all values of θ_c and thus cannot be used to motivate an error signal in a closed-loop configuration. In order to get a nonzero log-likelihood function, it can be shown that one must retain the first $M/2$ terms in the power series expansion. Thus, for example, for $M = 4$ and $M = 8$, we should use the approximations

$$\tanh x \cong \begin{cases} x - \frac{x^3}{3}, & M = 4 \\ x - \frac{x^3}{3} + \frac{2x^5}{15} - \frac{17x^7}{315}, & M = 8 \end{cases} \quad (8-44)$$

and in addition

⁵ For $M = 4$, i.e., QPSK, the nonlinearity is still a hyperbolic tangent function that now resides in both the I and Q arms.

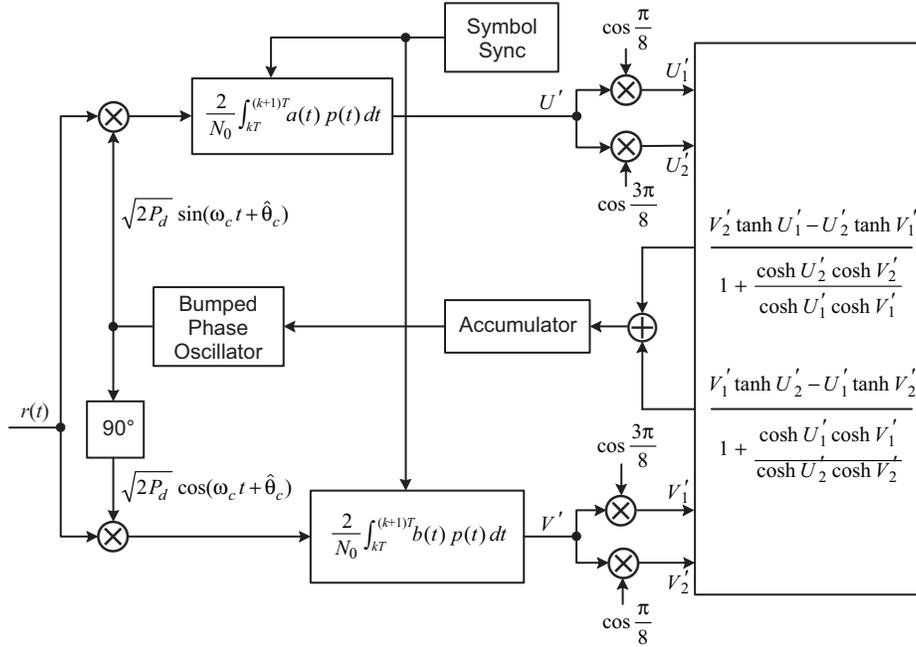


Fig. 8-9. A closed loop motivated by the MAP estimation of carrier phase for 8-PSK.

$$\cosh x \cong \begin{cases} 1 + \frac{x^2}{2}, & M = 4 \\ 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!}, & M = 8 \end{cases} \quad (8-45)$$

Applying these approximations and once again using passive arm filters, we obtain the single closed-loop structure illustrated in Fig. 8-12 capable of carrier synchronizing BPSK, QPSK, and 8-PSK. Several things are interesting about this structure. First of all, it is strictly of the I-Q type in that the loop error signal for all three modulations is derived from the I and Q arm filter outputs (i.e., U and V). The second and more interesting feature is that the error signal for the two higher-order modulations ($M = 4$ and $M = 8$) is derived from the multiplication of a *product* of two signals and a *difference of squares* of these same two signals. To see why this comes about, all one has to do is consider the following simple trigonometry.

For BPSK, the error signal is proportional to $\sin 2\phi$, which can be expressed as

$$\sin 2\phi = 2 \sin \phi \cos \phi \quad (8-46)$$

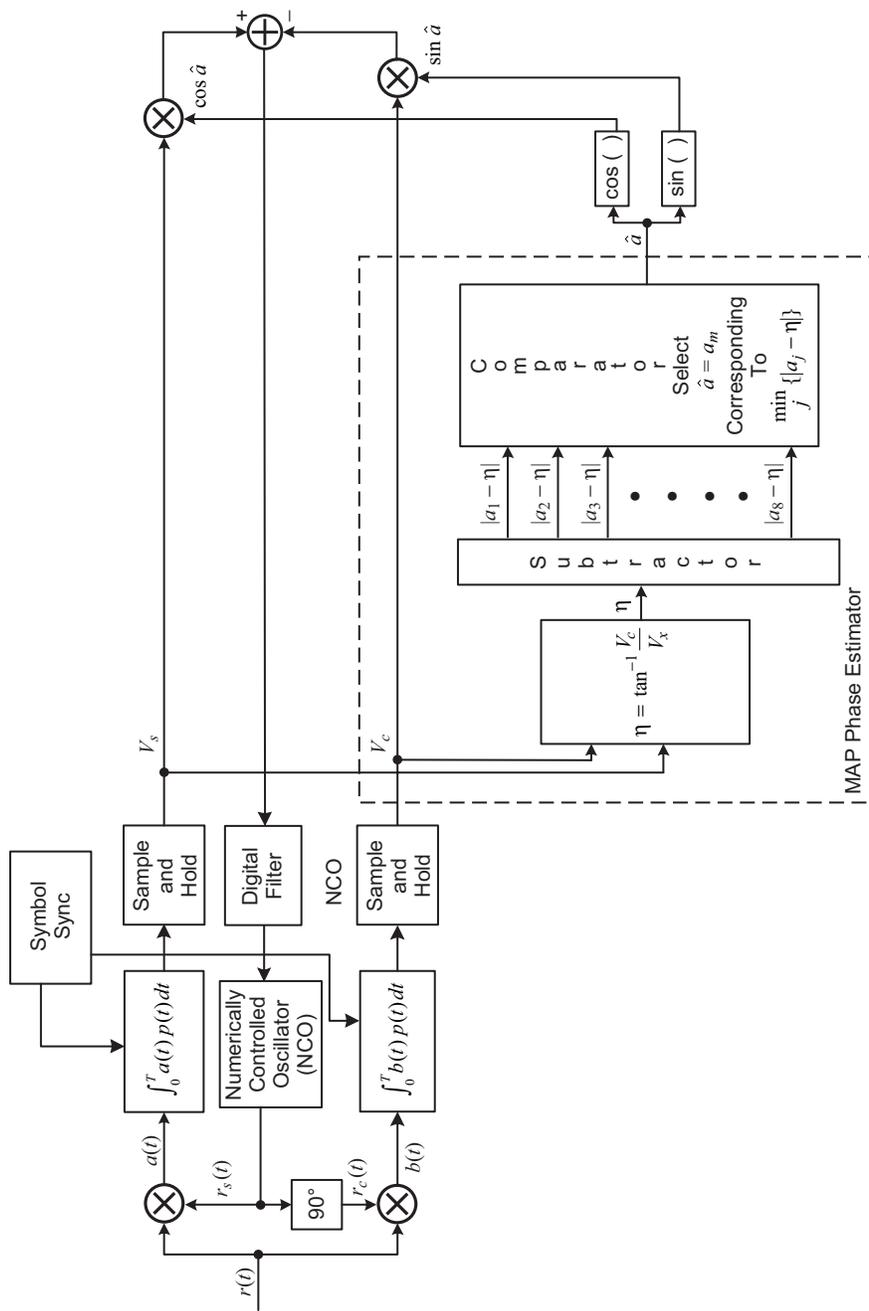


Fig. 8-10. High SNR approximation of the MAP estimation loop for 8-PSK.

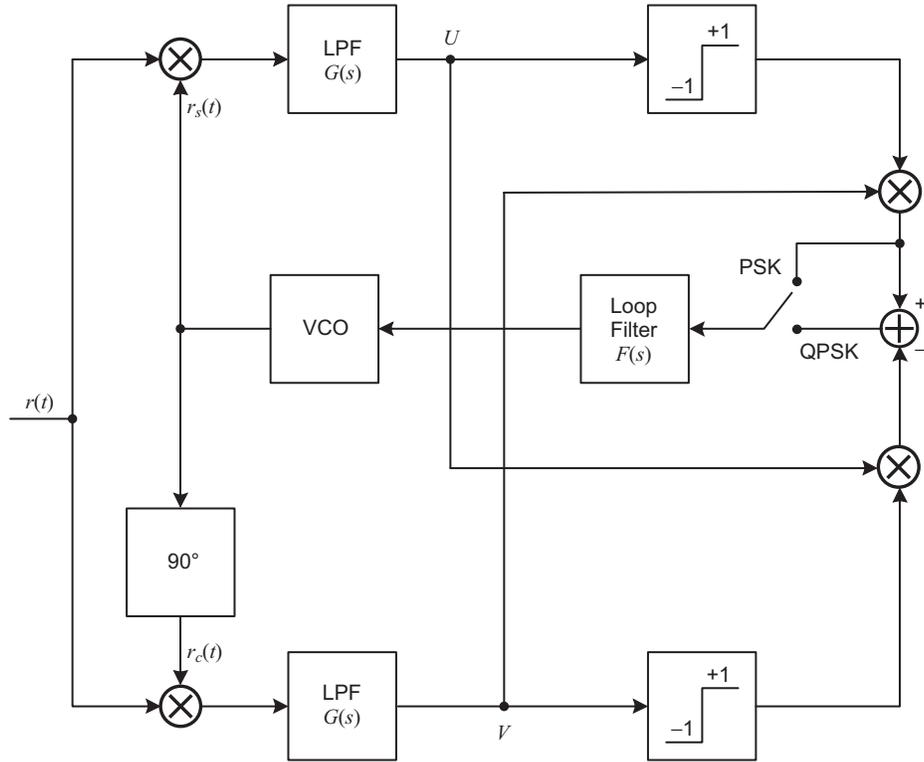


Fig. 8-11. High SNR approximation of the MAP estimation loop for BPSK and QPSK.

Thus, since U is proportional to $\sin \phi$ and V is proportional to $\cos \phi$, we see that the error signal is simply proportional to the product of U and V . For QPSK, the error signal is proportional to $\sin 4\phi$, which can be expressed as

$$\sin 4\phi = 2 \sin 2\phi \cos 2\phi = 2 \underbrace{\sin 2\phi}_{\text{error signal for BPSK}} \underbrace{(\cos^2 \phi - \sin^2 \phi)}_{\text{lock detector signal for BPSK}} \quad (8-47)$$

Similarly, for 8-PSK, the error signal is proportional to $\sin 4\phi$ which can be expressed as

$$\sin 8\phi = 2 \sin 4\phi \cos 4\phi = 2 \underbrace{\sin 4\phi}_{\text{error signal for QPSK}} \underbrace{(\cos^2 2\phi - \sin^2 2\phi)}_{\text{lock detector signal for QPSK}} \quad (8-48)$$

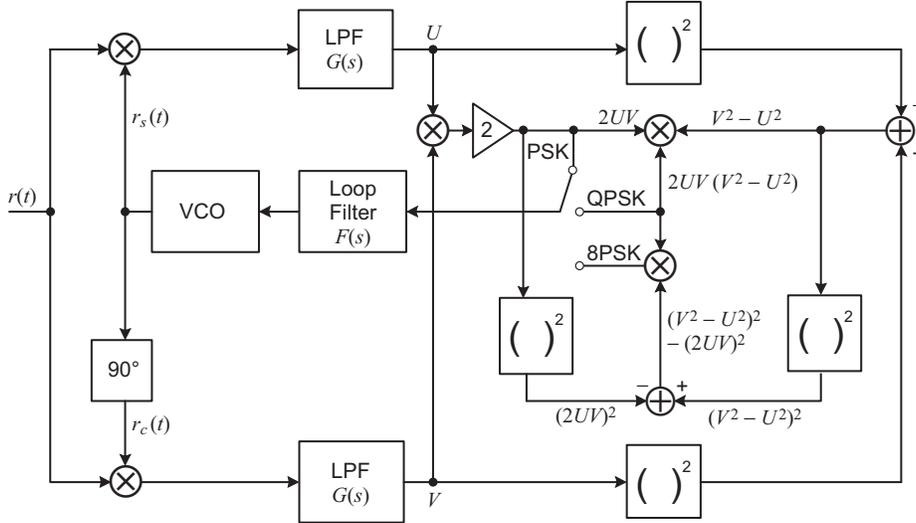


Fig. 8-12. A Costas-type loop capable of carrier tracking BPSK, QPSK, and 8-PSK; low SNR approximation of MAP estimation loop.

Examination of Fig. 8-12 clearly reveals that the error signals for the three modulations are formed in accordance with the relations in Eqs. (8-46) through (8-48). Thus, we conclude in general that the canonical structure for M -PSK has a front end (generation of the signals U and V) as in Fig. 8-12 and forms its error signal from the product of the error signal for $M/2$ -PSK and the lock detector signal for $M/2$ -PSK.

To use such a canonical configuration in an environment where the value of M is not known for certainty, one would proceed as follows. Since a BPSK loop cannot track QPSK, a QPSK cannot track 8-PSK, etc., one could start with the switch corresponding to the position of the lowest order modulation (BPSK) and sequentially move the switch to the positions of QPSK, 8-PSK, etc., until the loop locks. Another possibility would be to fix the switch in the position corresponding to the highest order loop and, provided that it would be capable of tracking all lower order modulations, accept the additional performance penalty incurred by using a higher order nonlinearity than necessary. To this end, as an example, we now examine the ability of a QPSK loop to track a BPSK signal. Since this issue appears not to be readily discussed in the literature, we shall be a bit more detailed here than we have been thus far in other parts of this chapter.

Consider the MAP estimation loop for QPSK illustrated in Fig. 8-13. The input to the loop is the BPSK signal $r(t) = s(t, \theta_c) + n(t)$, where

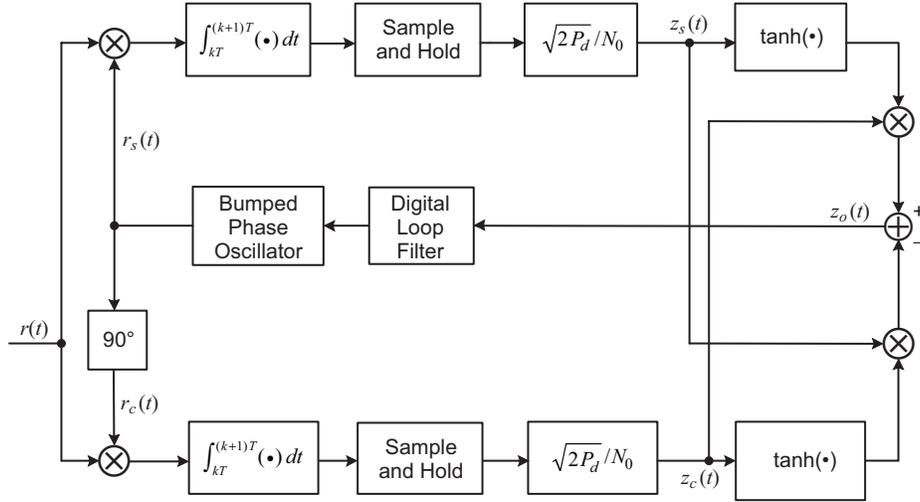


Fig. 8-13. The MAP estimation loop for carrier tracking QPSK with NRZ coding.

$$s(t, \theta_c) = \sqrt{2P_d}m(t) \sin(\omega_c t + \theta_c) \tag{8-49}$$

and the additive noise has the narrowband expansion

$$n(t) = \sqrt{2}[n_c(t) \cos(\omega_c t + \theta_c) - n_s(t) \sin(\omega_c t + \theta_c)] \tag{8-50}$$

where $n_c(t), n_s(t)$ are independent, low-pass Gaussian processes, each with single-sided PSD N_0 W/Hz and bandwidth $B_H < \omega_c/2\pi$. Assuming unit input I and Q phase detector (multiplier) gains and demodulation reference signals $r_c(t) = \sqrt{2} \cos(\omega_c t + \hat{\theta}_c)$ and $r_s(t) = \sqrt{2} \sin(\omega_c t + \hat{\theta}_c)$, then, after passing through the I and Q I&D filters of duration⁶ T and amplification by $\sqrt{2P_d}/N_0$, the sample-and-hold outputs $z_c(t)$ and $z_s(t)$ are given by

⁶ We remind the reader again that we are considering the case where the modulation bandwidth is held fixed and thus the I&D filters in all the configurations have a duration equal to the *symbol* time. Thus, while for the same *information* (bit) rate one would associate two BPSK bits with a QPSK symbol, for the same symbol rate, the I&D filters would correspond to a single bit interval for BPSK.

$$\begin{aligned}
z_s(t) &\triangleq \frac{\sqrt{2P_d}}{N_0} \int_{kT}^{(k+1)T} r(t)r_s(t)dt \\
&= \sqrt{2}R_dc_k \cos \phi - \sqrt{R_d}X_1 \cos \phi - \sqrt{R_d}X_2 \sin \phi \\
z_c(t) &\triangleq \frac{\sqrt{2P_d}}{N_0} \int_{kT}^{(k+1)T} r(t)r_c(t)dt \\
&= \sqrt{2}R_dc_k \sin \phi - \sqrt{R_d}X_1 \sin \phi + \sqrt{R_d}X_2 \cos \phi
\end{aligned} \tag{8-51}$$

where $(k+1)T \leq t \leq (k+2)T$ and where X_1, X_2 are zero-mean, unit variance independent Gaussian RVs. Multiplying $z_c(t)$ by the nonlinearly processed $z_s(t)$ and vice versa gives the dynamic error signal

$$z_o(t) = -z_s(t) \tanh z_c(t) + z_c(t) \tanh z_s(t) \tag{8-52}$$

As in all analyses of this type, the tracking performance of a loop can, in its linear region of operation (small phase error), be determined by examining the equivalent signal and noise components of the $z_o(t)$ process, more specifically, the slope of the equivalent S-curve at $\phi = 0$ and the variance of the equivalent additive noise. This makes the usual assumption that the loop bandwidth is much less than the data bandwidth.

Since X_1 and X_2 are zero-mean random variables, then, from Eq. (8-52) together with Eq. (8-51), the signal component of $z_o(t)$ has a mean, i.e., the S-curve of the loop, given by

$$\begin{aligned}
\eta(\phi) &= -4\sqrt{2}R_d(c_k \cos \phi) \tanh \left[\sqrt{2}R_dc_k \sin \phi \right. \\
&\quad \left. - \sqrt{R_d}X_1 \sin \phi + \sqrt{R_d}X_2 \cos \phi \right]^{X_1, X_2, c_k} \\
&\quad + 4\sqrt{2}R_d(c_k \sin \phi) \tanh \left[\sqrt{2}R_dc_k \cos \phi \right. \\
&\quad \left. - \sqrt{R_d}X_1 \cos \phi - \sqrt{R_d}X_2 \sin \phi \right]^{X_1, X_2, c_k} \\
&= -4\sqrt{2}R_d \cos \phi \tanh \left(\sqrt{2}R_d \sin \phi - \sqrt{R_d}X_1 \sin \phi + \sqrt{R_d}X_2 \cos \phi \right)^{X_1, X_2} \\
&\quad + 4\sqrt{2}R_d \sin \phi \tanh \left(\sqrt{2}R_d \cos \phi - \sqrt{R_d}X_1 \cos \phi - \sqrt{R_d}X_2 \sin \phi \right)^{X_1, X_2}
\end{aligned} \tag{8-53}$$

This S-curve is an odd function of ϕ and is periodic with period $\pi/2$, which implies a phase ambiguity for the loop of 90 deg. This ambiguity, which must be resolved for successful data detection, is the same as would be the case for the loop tracking a QPSK signal. Aside from the ambiguity itself, it can also be observed from Eq. (8-53) that the S-curve passes through zero midway between these potential lock points, namely, at $\phi = \pi/4$. Thus, in order to determine whether the loop will correctly lock at $\phi = 0$ (assuming resolution of the ambiguity) or incorrectly at $\phi = \pi/4$, one needs to show that the slope of the S-curve at the former phase error value is positive (which would imply that the slope at the latter value is negative).

The slope of the S-curve at $\phi = 0$ is obtained by differentiating Eq. (8-53) with respect to 4ϕ and evaluating the result at this same phase error value. Recognizing that $\overline{X \operatorname{sech}^2 \sqrt{R_d X}} = 0$, it is straightforward to show that

$$K_\eta = \left. \frac{d\eta(\phi)}{d(4\phi)} \right|_{\phi=0} = \sqrt{2R_d} \overline{\tanh(\sqrt{2R_d} - \sqrt{R_d X})^X} - 2R_d^2 \overline{\operatorname{sech}^2(\sqrt{R_d X})^X} \quad (8-54)$$

If we now make the low SNR approximation of the nonlinearities in Eq. (8-54) using only the first terms of their Taylor series expansions, i.e.,

$$\tanh x \cong x, \quad \operatorname{sech}^2 x \cong 1 \quad (8-55)$$

then since the X 's are zero mean, we immediately get

$$K_\eta = 0 \quad (8-56)$$

which implies that the loop would be unable to lock at all. Thus, as was the case in deriving the QPSK MAP estimation loop from maximum-likelihood considerations, we must include the next terms in the Taylor series expansions of the nonlinearities. That is, we apply

$$\tanh x \cong x - \frac{x^3}{3}, \quad \operatorname{sech}^2 x = \frac{d}{dx} \tanh x \cong 1 - x^2 \quad (8-57)$$

which results in the QPSK portion of the implementation in Fig. 8-12 but with I&D arm filters. When this is done, making use of the moments of a Gaussian RV, we obtain

$$\overline{\tanh(a - bX)}^X \cong a \left(1 - \frac{a^2}{3} - b^2\right) \quad (8-58)$$

$$\overline{\operatorname{sech}^2 bX}^X \cong 1 - b^2$$

and thus

$$\overline{\tanh(a - bX)}^X - a \overline{\operatorname{sech}^2 bX}^X \cong -\frac{a^3}{3} \quad (8-59)$$

Using Eq. (8-59) in Eq. (8-54) gives

$$K_\eta = \sqrt{2}R_d \left(-\frac{1}{3} \left(\sqrt{2}R_d\right)^3\right) = -\frac{4}{3}R_d^4 \quad (8-60)$$

which is negative and thereby would require that the signs on the summer at the input to the accumulator in Fig. 8-13 be reversed, i.e., the polarity of the error signal flipped, in order for the loop to correctly lock at $\phi = 0$ (and equally well at the ambiguity phase error values $\phi = \pi/2, \pi, 3\pi/2$).

Proceeding now to an evaluation of the equivalent noise PSD, ignoring the self-noise of the signal component, the noise component of $z_o(t)$ (evaluated at $\phi = 0$) is

$$N_e(t) = \sqrt{R_d}X_1 \tanh(\sqrt{R_d}X_2) + \sqrt{R_d}X_2 \tanh[\sqrt{2}R_d c_k - \sqrt{R_d}X_1] \quad (8-61)$$

which has zero mean and variance

$$\sigma_{N_e}^2 = R_d \left\{ \overline{\tanh^2(\sqrt{R_d}X)}^X + \overline{\tanh^2(\sqrt{2}R_d - \sqrt{R_d}X)}^X \right. \\ \left. + 2X \overline{\tanh(\sqrt{R_d}X)}^X \left[\overline{X \tanh(\sqrt{2}R_d - \sqrt{R_d}X)}^X \right] \right\} \quad (8-62)$$

Once again we apply the approximations of the nonlinearities in Eq. (8-57) to evaluate the variance in Eq. (8-62). In particular, the following results for each statistical average are obtained:

$$\begin{aligned}
\overline{\tanh^2(\sqrt{2}R_d - \sqrt{R_d}X)}^X &= \frac{8}{9}R_d^6 + \frac{20}{3}R_d^5 + \frac{22}{3}R_d^4 - \frac{19}{3}R_d^3 + R_d \\
\overline{\tanh^2(\sqrt{R_d}X)}^X &= R_d - 2R_d^2 + \frac{5}{3}R_d^3 \\
X \overline{\tanh(\sqrt{R_d}X)}^X &= \sqrt{R_d}(1 - R_d) \\
X \overline{\tanh(\sqrt{2}R_d - \sqrt{R_d}X)}^X &= -\sqrt{R_d}(1 - R_d - 2R_d^2)
\end{aligned} \tag{8-63}$$

Substituting the results of Eq. (8-63) into Eq. (8-62) results after some simplification in

$$\sigma_{N_e}^2 = R_d \left(\frac{8}{9}R_d^6 + \frac{20}{3}R_d^5 + \frac{10}{3}R_d^4 - \frac{8}{3}R_d^3 + 2R_d^2 \right) \tag{8-64}$$

Because of the I&D arm filters in Fig. 8-13, the noise process of Eq. (8-61) is piecewise constant over intervals of T -seconds duration. Thus, as long as the loop bandwidth is much less than the data bandwidth, this process can be approximated, as has been done in the past, by a delta-correlated process with correlation function given by

$$R_{N_e}(\tau) \triangleq \overline{N_e(t)N_e(t+\tau)} = \begin{cases} \sigma_{N_e}^2 \left[1 - \frac{|\tau|}{T} \right], & |\tau| \leq T \\ 0; & |\tau| > T \end{cases} \tag{8-65}$$

with equivalent single-sided noise spectral density

$$N'_0 \triangleq 2 \int_{-\infty}^{\infty} R_{N_e}(\tau) d\tau = 2\sigma_{N_e}^2 T \tag{8-66}$$

As such, the linearized phase error variance is given by

$$\sigma_\phi^2 = N'_0 B_L / K_\eta^2 = (\rho S_L)^{-1} \tag{8-67}$$

where ρ is the linear loop (PLL) SNR as defined in Eq. (8-10) and S_L is the "quadrupling loss" which reflects the penalty paid due to the signal and noise

cross-products present in $z_o(t)$. Substituting Eq. (8-66) in Eq. (8-67), the quadrupling loss can be identified as

$$S_L = \left(\frac{1}{2R_d} \right) \frac{K_\eta^2}{\sigma_{N_e}^2} = \frac{K_\eta^2 / (2R_d^2)}{\sigma_{N_e}^2 / R_d} \quad (8-68)$$

Finally, substituting Eqs. (8-60) and (8-64) in Eq. (8-68) gives the desired result:

$$S_L = \frac{\frac{8}{9}R_d^6}{\frac{8}{9}R_d^6 + \frac{20}{3}R_d^5 + \frac{10}{3}R_d^4 - \frac{8}{3}R_d^3 + 2R_d^2} = \frac{1}{1 + \frac{15}{2R_d} + \frac{15}{4R_d^2} - \frac{3}{R_d^3} + \frac{9}{4R_d^4}} \quad (8-69)$$

It is interesting to compare this loss to that which would result from the same loop tracking a QPSK signal. In particular, for the MAP estimation loop with a QPSK input, the squaring loss is given by [4]

$$S_L = \frac{\left[\frac{\tanh(R_d - \sqrt{R_d}X)^X - R_d \operatorname{sech}^2(\sqrt{R_d}X)^X}{(1 + R_d) \tanh^2(R_d - \sqrt{R_d}X)^X} - \left[\frac{X \tanh(R_d - \sqrt{R_d}X)^X - \sqrt{R_d} \tanh(R_d - \sqrt{R_d}X)}{1 + R_d} \right]^2 \right]}{\left[\frac{\tanh(R_d - \sqrt{R_d}X)^X - R_d \operatorname{sech}^2(\sqrt{R_d}X)^X}{(1 + R_d) \tanh^2(R_d - \sqrt{R_d}X)^X} - \left[\frac{X \tanh(R_d - \sqrt{R_d}X)^X - \sqrt{R_d} \tanh(R_d - \sqrt{R_d}X)}{1 + R_d} \right]^2 \right]} \quad (8-70)$$

which for the low SNR approximation loop (the QPSK portion of Fig. 8-12 with I&D arm filters) reduces, after considerable manipulation, to [6]

$$S_L = \frac{1}{1 + \frac{9}{2R_d} + \frac{6}{R_d^2} + \frac{3}{2R_d^3}} \quad (8-71)$$

Thus, from a comparison of Eqs. (8-69) and (8-71), we conclude that while the QPSK carrier tracking loop is capable of tracking a BPSK signal it does so with a different mean-squared tracking error performance than for a QPSK input signal. Furthermore, the quadrupling loss of Eq. (8-69) exceeds the squaring loss of Eq. (8-13) for all SNRs. The more important issue, however, is the means by which the 90-deg phase ambiguity must be resolved. Whereas for a BPSK loop tracking a BPSK signal one can easily resolve the associated 180-deg phase

ambiguity by differentially encoding the binary data, resolving the 90-deg phase ambiguity associated with the QPSK loop cannot be resolved solely by the same means. In the case of the latter, one must in addition detect the data from the outputs of *both* the I and Q channels and choose the one that has the higher reliability.

In view of the issues brought forth in the above example, it appears that the preferred solution for autonomous operation is not to use a single loop for all modulation orders but rather to first classify the modulation, i.e., determine its order and then place the switch in Fig. 8-12 in its appropriate position.

8.4.2 QAM and Unbalanced QPSK

It is straightforward to extend the notions described above to other modulations with a quadrature structure such as QAM and unbalanced quadrature phase-shift keying (UQPSK) [6]. For example, for square QAM with $M = K^2$ symbols described by

$$s(t, \theta_c) = \sqrt{2}Am_I(t) \cos(\omega_c t + \theta_c) + \sqrt{2}Am_Q(t) \sin(\omega_c t + \theta_c) \quad (8-72)$$

$$A = \sqrt{\frac{3}{2(M-1)}} P_d$$

where $m_I(t), m_Q(t)$ are the quadrature data modulations of rate $1/T$ taking on values $\pm 1, \pm 3, \dots, \pm\sqrt{M} - 1$, the derivative of the log-likelihood function becomes

$$\frac{d\Lambda(\theta_c)}{d\theta_c} \simeq \sum_{k=0}^{K-1} \frac{\sum_{l=1}^{\sqrt{M}/2} \exp(-c_l^2 R_d) c_l U' \sinh(c_l U')}{\sum_{l=1}^{\sqrt{M}/2} \exp(-c_l^2 R_d) \cosh(c_l U')} - \sum_{k=0}^{K-1} \frac{\sum_{l=1}^{\sqrt{M}/2} \exp(-c_l^2 R_d) c_l U' \sinh(c_l V')}{\sum_{l=1}^{\sqrt{M}/2} \exp(-c_l^2 R_d) \cosh(c_l V')} \quad (8-73)$$

where $c_l = 2l - 1$, $U' = \sqrt{3/(M-1)}U$, $V' = \sqrt{3/(M-1)}V$, and, as before, $R_d = P_d T / N_0$ is the data SNR. A closed-loop carrier synchronizer motivated by this MAP estimation approach is illustrated in Fig. 8-14. Here again we

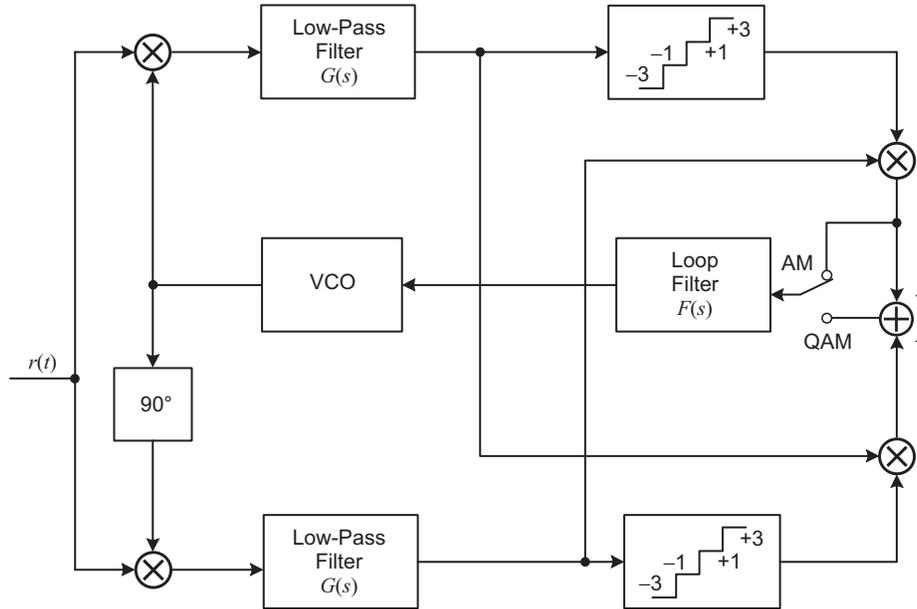


Fig. 8-15. High SNR approximation of the MAP estimation loop for amplitude modulation (AM) and QAM.

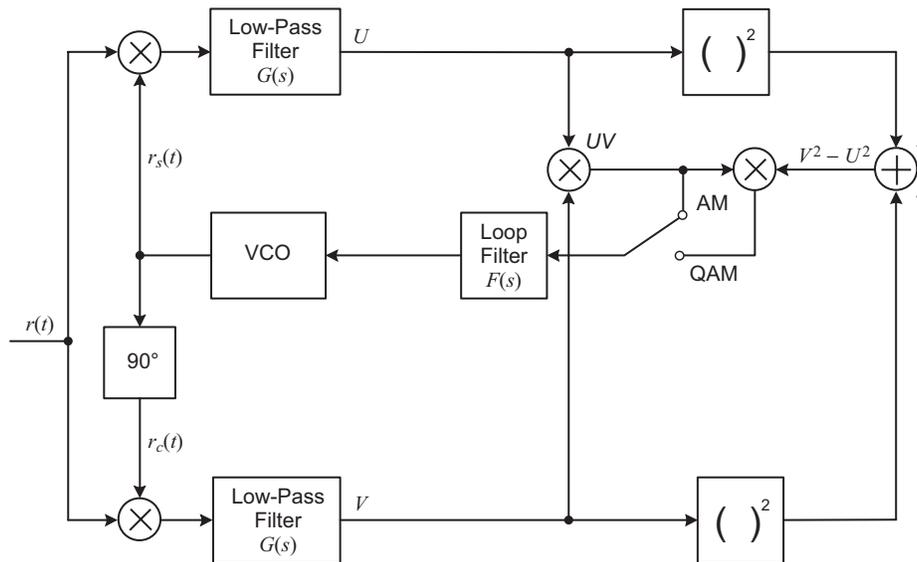


Fig. 8-16. Low SNR approximation for the MAP estimation loop for AM and QAM (the same as the low SNR approximation for PSK and QPSK).

$$\begin{aligned}
\frac{d\Lambda(\theta_c)}{d\theta_c} &\cong \sum_{k=0}^{K_2} \left(\frac{2\sqrt{2P_2}}{N_0} \int_{\delta_k}^{\delta_{k+1}} r(t)p_2(t - \delta_k) \cos(\omega_c t + \theta_c) dt \right) \\
&\quad \times \tanh \left\{ \frac{2\sqrt{2P_2}}{N_0} \int_{\delta_k}^{\delta_{k+1}} r(t)p_2(t - \delta_k) \sin(\omega_c t + \theta_c) dt \right\} \\
&\quad - \sum_{k=0}^{K_1} \left(\frac{2\sqrt{2P_1}}{N_0} \int_{\tau_k}^{\tau_{k+1}} r(t)p_1(t - \tau_k) \cos(\omega_c t + \theta_c) dt \right) \\
&\quad \times \tanh \left\{ \frac{2\sqrt{2P_1}}{N_0} \int_{\tau_k}^{\tau_{k+1}} r(t)p_1(t - \tau_k) \sin(\omega_c t + \theta_c) dt \right\} \quad (8-75)
\end{aligned}$$

In Eq. (8-75), $\tau_k (k = 0, 1, \dots, K_1)$ is the ordered set of time instants at which the modulation $m_1(t)$ may potentially have a symbol transition in the observation interval ($0 \leq t \leq T_o$). Similarly, $\delta_k (k = 0, 1, \dots, K_2)$ is the ordered set of time instants at which the modulation $m_2(t)$ may potentially have a symbol transition in the same observation interval. Note that, since we have not restricted T_1 or T_2 to be integer related, it is possible that the observation may not contain an integer number of symbol intervals of one of the two modulations. For this reason, we allow the summations in Eq. (8-75) to extend over $K + 1$ symbols. Furthermore, no restriction is placed on the relative synchronization between the taus and the deltas.

Figure 8-17 illustrates the MAP estimation closed loop that results from using Eq. (8-75) as an error signal. As before, one can use the approximations of the hyperbolic tangent nonlinearity as given in Eq. (8-14) to produce low and high SNR configurations. The difficulty with using the small argument approximation is that, as the ratio of rates and powers both approach unity, i.e., balanced QPSK, the two pairs of matched filters (or equivalently the two pairs of arm filters in the passive implementation) become identical, and thus the error signal at the input to the loop filter goes to zero for all phase errors. Thus, as was done previously in deriving the MAP estimation loop for balanced QPSK, one must consider the first two terms of the power series expansion of the hyperbolic tangent nonlinearity as in Eq. (8-44), which results in the configuration illustrated in Fig. 8-18 (assuming the passive arm filter implementation). Note that this two-channel Costas loop reduces (except for the 1/3 gain factor) to Fig. 8-16 when the transmitted signal becomes balanced QPSK. Thus, this configuration is capable of tracking unbalanced as well as balanced QPSK.

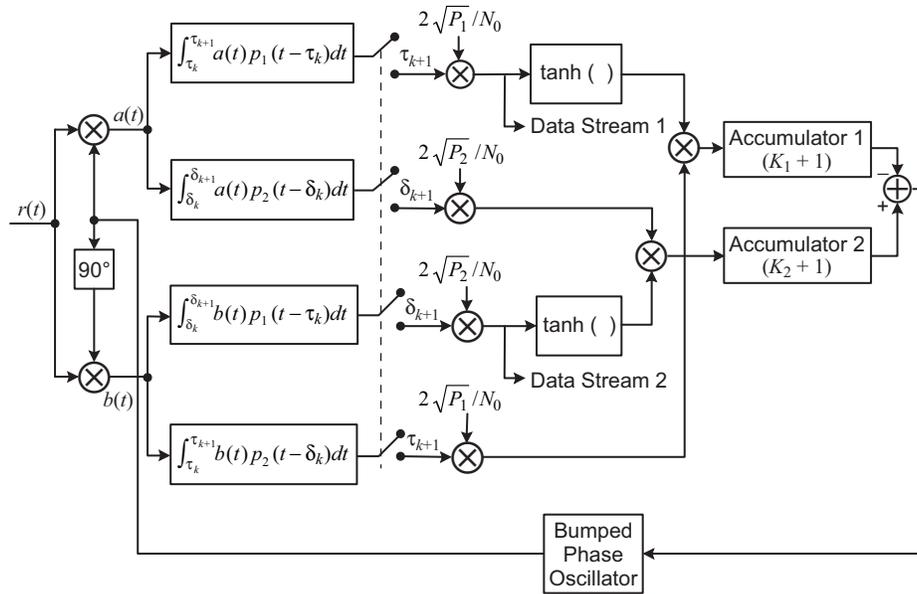


Fig. 8-17. A closed loop motivated by the MAP estimation of carrier phase for unbalanced QPSK.

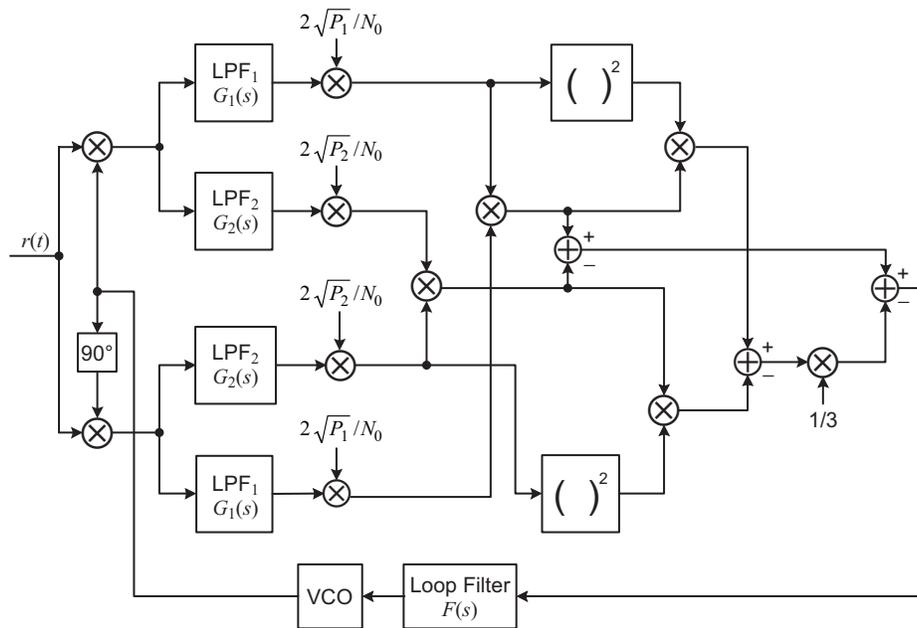


Fig. 8-18. Low SNR approximation for the MAP estimation loop for unbalanced QPSK.

We conclude by noting that, depending on the ratio of powers in the two channel of the UQPSK signal, it is possible to employ just the simple biphasic or quadriphase Costas loops previously discussed, bearing in mind, however, that the performance of such would then be inferior to that of the loop in Fig. 8-18.

8.4.3 $\pi/4$ Differentially Encoded QPSK

As a final modulation form, we consider the case of $\pi/4$ differentially encoded QPSK in which the information phase symbols are chosen from the set $(\pm\pi/4, \pm3\pi/4)$ and are differentially encoded prior to transmission. Denoting the information symbol in the n th symbol interval by $\Delta\phi_n$, then the actual transmitted symbol in the same interval is given by $\phi_n = \phi_{n-1} + \Delta\phi_n$, which, in view of the set used to define $\Delta\phi_n$ as given above, alternates between the allowable sets $(0, \pi/2, \pi, 3\pi/2)$ and $(\pm\pi/4, \pm3\pi/4)$ in successive transmission intervals. Because of this $\pi/4$ rad rotation of the transmitted signaling constellation from symbol to symbol, the maximum instantaneous phase change between two successive symbols is $3\pi/4$ rad. This is to be compared to a maximum instantaneous phase change of π rad for the case where the signaling constellation does not rotate from symbol to symbol or, equivalently, the information symbols to be differentially encoded are chosen from the set $(0, \pi/2, \pi, 3\pi/2)$. Reducing the maximum phase jump from π to $3\pi/4$ reduces the envelope fluctuation in the signal, which is desirable on nonlinear channels to prevent spectral side lobes from being regenerated after having been filtered.

Applying the same MAP phase estimation approach as previously used to motivate a closed-loop structure, it can be shown that the appropriate carrier synchronizer for this so-called $\pi/4$ differentially encoded QPSK modulation is as illustrated in Fig. 8-19. Once again, if desired, one can apply the appropriate approximations to the hyperbolic tangent function to arrive at low and high SNR implementations of this generic structure.

References

- [1] M. K. Simon, "Optimum Receiver Structures for Phase-Multiplexed Modulations," *IEEE Transactions on Communications*, vol. COM-26, pp. 865–872, June 1978.
- [2] M. K. Simon, "Further Results on Optimum Receiver Structures for Digital Amplitude and Phase Modulated Signals," *International Conference on Communications (ICC'78) Record*, Toronto, Ontario, Canada, pp. 42.1.1–42.1.7, June 1978.

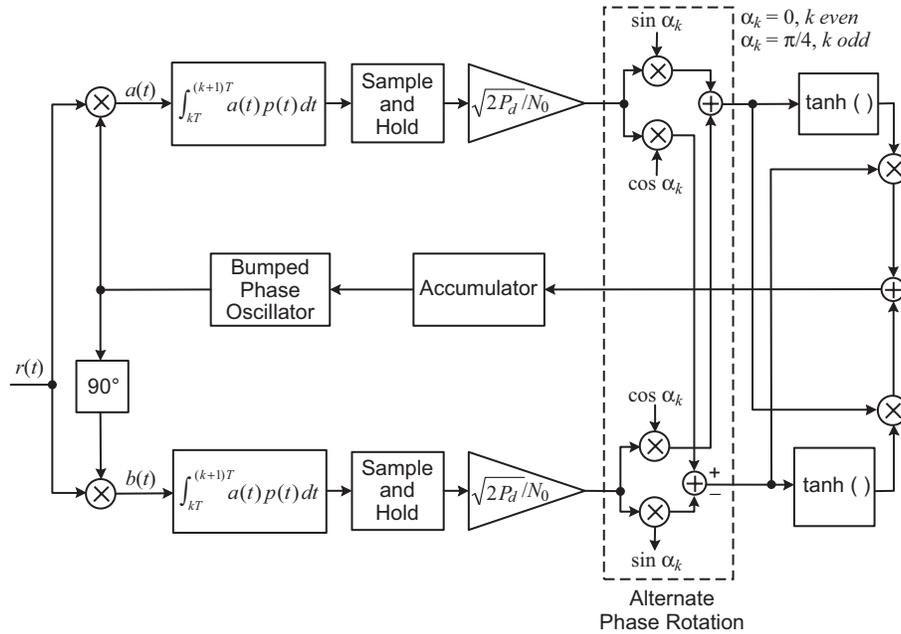


Fig. 8-19. MAP estimation loop for carrier synchronization of $\pi/4$ differentially encoded QPSK.

- [3] P. W. Kinman, "TLM-21 DSN Telemetry System, Block-V Receiver," *DSMS Telecommunications Link Design Handbook*, 810-5, Rev. D, Jet Propulsion Laboratory, Pasadena, California, December 1996.
- [4] "Proximity-1 Space Link Protocol," CCSDS 211.0-B-1, *Blue Book*, Issue 1, October 2002.
- [5] M. K. Simon and S. Butman, "On the Receiver Structure for a Single-Channel, Phase-Coherent Communication System," *Space Program Summary 37-62*, vol. III, Jet Propulsion Laboratory, Pasadena, California, pp. 103–108, April 1970.
- [6] M. K. Simon, "On the Optimality of the MAP Estimation Loop for Tracking BPSK and QPSK Signals," *IEEE Transactions on Communications*, vol. COM-27, pp. 158–165, January 1979.
- [7] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part III*, New York: John Wiley, 1968–1971.

- [8] C. R. Cahn, “Improving Frequency Acquisition of a Costas Loop,” *IEEE Transactions on Communications*, vol. COM-25, pp. 1453–1459, December 1977.
- [9] M. K. Simon and W. C. Lindsey, “Optimum Performance of Suppressed Carrier Receivers with Costas Loop Tracking,” *IEEE Transactions on Communications*, vol. COM-25, pp. 215–227, February 1977.
- [10] M. K. Simon, “On the Calculation of Squaring Loss in Costas Loops with Arbitrary Arm Filters,” *IEEE Transactions on Communications*, vol. COM-26, pp. 179–184, January 1978.
- [11] M. K. Simon, “The Effects of Residual Carrier on Costas Loop Performance as Applied to the Shuttle S-Band Uplink,” *IEEE Transactions on Communications*, vol. COM-26, pp. 1542–1548, November 1978.

Appendix 8-A

Cramer–Rao Bound on the Variance of the Error in Estimating the Carrier Phase of a BPSK Signal

As we have seen earlier in the chapter, MAP estimation of the carrier phase of a BPSK-modulated carrier can be formulated as the solution to a problem in which a suitably defined likelihood function is maximized with respect to the carrier phase parameter. As a check on the efficiency of the estimator so obtained, one often derives the Cramer–Rao lower bound [7] on the performance measure being considered. Here we derive the Cramer–Rao bound on the variance of the error in the MAP estimation of the phase of a BPSK signal. More often than not, what is typically done in the literature in such applications is to use the result obtained for an unmodulated carrier, which as we shall see is a proper thing to do only at high SNR.

Consider an observation over an interval $T_o = KT$ seconds of a BPSK-modulated carrier in AWGN, where K denotes the number of bits in the observation and T is the bit time (the reciprocal of the bit rate). The received signal in the k th bit time interval $kT \leq t \leq (k+1)T$ takes the form

$$r(t) = s(t, \theta_c) + n(t) = \sqrt{2P_d} c_k \cos(\omega_c t + \theta_c) + n(t) \quad (\text{A-1})$$

where P_d is the received data power, ω_c is the radian carrier frequency, c_k is the k th bit taking on equiprobable ± 1 values, and $n(t)$ is the AWGN with single-sided power spectral density N_0 W/Hz. The likelihood function for the observation $r(t)$ conditioned on the data sequence $\mathbf{c} = (c_1, c_2, \dots, c_K)$ and the carrier phase θ_c is well-known to be

$$\begin{aligned} p(r(t) | \theta_c, \mathbf{c}) &= C \exp \left\{ -\frac{1}{N_0} \int_0^{T_o} [r(t) - s(t, \theta_c)]^2 dt \right\} \\ &= C \prod_{k=0}^{K-1} \exp \left\{ -\frac{1}{N_0} \int_{kT}^{(k+1)T} [r(t) - s(t, \theta_c)]^2 dt \right\} \end{aligned} \quad (\text{A-2})$$

To determine the Cramer-Rao bound, it is necessary to evaluate the parameter

$$\gamma \triangleq E \left\{ \left(\frac{\partial \ln p(r(t) | \theta_c)}{\partial \theta_c} \right)^2 \right\} \quad (\text{A-3})$$

Thus, we must first average Eq. (A-2) over the bit sequence in order to arrive at $p(r(t) | \theta_c)$. Denoting $s_1(t, \theta_c)$ as the transmitted signal in the k th bit interval when $c_k = 1$ and likewise $s_{-1}(t, \theta_c)$ as the transmitted signal in the k th bit interval when $c_k = -1$, then

$$\begin{aligned} p(r(t) | \theta_c) &= C \prod_{k=0}^{K-1} \left[\frac{1}{2} \exp \left\{ -\frac{1}{N_0} \int_{(k-1)T}^{kT} [r(t) - s_1(t, \theta_c)]^2 dt \right\} \right. \\ &\quad \left. + \frac{1}{2} \exp \left\{ -\frac{1}{N_0} \int_{kT}^{(k+1)T} [r(t) - s_{-1}(t, \theta_c)]^2 dt \right\} \right] \\ &= C \prod_{k=0}^{K-1} \exp \left\{ -\frac{1}{N_0} \int_{kT}^{(k+1)T} r^2(t) dt \right\} \exp(-R_d) \\ &\quad \times \cosh \left\{ \frac{2}{N_0} \int_{kT}^{(k+1)T} r(t) s_1(t, \theta_c) dt \right\} \end{aligned} \quad (\text{A-4})$$

where $R_d = P_d T / N_0$ is, as before, the data SNR. Taking the natural logarithm of Eq. (A-4) and differentiating with respect to θ_c gives

$$\begin{aligned}
\frac{\partial \ln p(r(t) | \theta_c)}{\partial \theta_c} &= \sum_{k=0}^{K-1} \frac{\partial}{\partial \theta_c} \ln \cosh \left\{ \frac{2}{N_0} \int_{kT}^{(k+1)T} r(t) s_1(t, \theta_c) dt \right\} \\
&= \sum_{k=0}^{K-1} \tanh \left\{ \frac{2}{N_0} \int_{kT}^{(k+1)T} r(t) s_1(t, \theta_c) dt \right\} \\
&\quad \times \left[\left\{ \frac{2}{N_0} \int_{kT}^{(k+1)T} r(t) \frac{\partial s_1(t, \theta_c)}{\partial \theta_c} dt \right\} \right] \\
&= - \sum_{k=0}^{K-1} \tanh \left\{ \frac{2\sqrt{2P_d}}{N_0} \int_{kT}^{(k+1)T} r(t) \cos(\omega_c t + \theta_c) dt \right\} \\
&\quad \times \left[\frac{2\sqrt{2P_d}}{N_0} \int_{kT}^{(k+1)T} r(t) \sin(\omega_c t + \theta_c) dt \right] \tag{A-5}
\end{aligned}$$

Consider first the evaluation of Eq. (A-3) together with Eq. (A-5) for the case of high SNR, where the tanh nonlinearity is approximated by the signum function. Letting

$$\begin{aligned}
I_k &= \frac{2\sqrt{2P_d}}{N_0} \int_{kT}^{(k+1)T} r(t) \cos(\omega_c t + \theta_c) dt \\
Q_k &= \frac{2\sqrt{2P_d}}{N_0} \int_{kT}^{(k+1)T} r(t) \sin(\omega_c t + \theta_c) dt
\end{aligned} \tag{A-6}$$

then for high SNR we have

$$\frac{\partial \ln p(y(t) | \theta_c)}{\partial \theta_c} = - \sum_{k=0}^{K-1} Q_k \operatorname{sgn} I_k \tag{A-7}$$

and because the I_k 's and Q_k 's are iid and independent of each other,

$$\begin{aligned}
E \left\{ \left(\frac{\partial \ln p(r(t) | \theta_c)}{\partial \theta_c} \right)^2 \right\} &= \sum_{k=0}^{K-1} E \left\{ Q_k^2 \overbrace{\text{sgn}^2 I_k}^{=1} \right\} \\
&+ 2 \sum_{\substack{k=0 \\ k \neq l}}^{K-1} \sum_{l=0}^{K-1} E \{ Q_k \text{sgn} I_k \} E \{ Q_l \text{sgn} I_l \} \quad (\text{A-8})
\end{aligned}$$

Assuming that indeed $s_1(t, \theta_c)$ was transmitted in the k th interval, i.e., $r(t) = s_1(t, \theta_c) + n(t)$, then substituting Eq. (A-6) into Eq. (A-8) and carrying out the expectation over the noise gives, after simplification,

$$\begin{aligned}
E \{ Q_k^2 \} &= 2R_d \\
E \{ Q_k \text{sgn} I_k \} &= 0
\end{aligned} \quad (\text{A-9})$$

If instead one assumes that the transmitted signal was $s_{-1}(t, \theta_c)$ rather than $s_1(t, \theta_c)$, then one arrives at the identical result as Eq. (A-9). Thus, independent of the actual transmitted data sequence, we have

$$E \left\{ \left(\frac{\partial \ln p(r(t) | \theta_c)}{\partial \theta_c} \right)^2 \right\} = K (2R_d) \quad (\text{A-10})$$

Finally, the Cramer–Rao bound on the variance of the unbiased estimation error $\phi \triangleq \theta_c - \hat{\theta}_c$ is given by

$$\sigma_\phi^2 \geq \left[E \left\{ \left(\frac{\partial \ln p(r(t) | \theta_c)}{\partial \theta_c} \right)^2 \right\} \right]^{-1} = \frac{1}{K (2R_d)} \quad (\text{A-11})$$

For an unmodulated carrier of energy $E = P_d T_o$ over the observation, the Cramer–Rao bound on the variance of the estimation error is given by

$$\sigma_\phi^2 \geq \frac{1}{2E/N_0} \quad (\text{A-12})$$

which in view of the relation $T_o = KT$ is identical to Eq. (A-8). Thus, as previously mentioned, we see that the Cramer–Rao bound for the modulated carrier is equivalent to that for the unmodulated carrier at high SNR.

For low SNR, one approximates the tanh nonlinearity by a linear function, i.e., $\tanh x = x$. Thus, the analogous relation to Eq. (A-7) is now

$$\frac{\partial \ln p(r(t)|\theta_c)}{\partial \theta_c} = - \sum_{k=1}^K Q_k I_k \quad (\text{A-13})$$

Again because of the independence of the I_k 's and Q_k 's we get

$$E \left\{ \left(\frac{\partial \ln p(r(t)|\theta_c)}{\partial \theta_c} \right)^2 \right\} = \sum_{k=1}^K E \{ Q_k^2 I_k^2 \} + 2 \sum_{\substack{k=1 \\ k \neq l}}^K \sum_{l=1}^K E \{ Q_k I_k \} E \{ Q_l I_l \} \quad (\text{A-14})$$

with (after much simplification)

$$E \{ Q_k^2 I_k^2 \} = (2R_d)^2 (1 + 2R_d) \quad (\text{A-15})$$

$$E \{ Q_k I_k \} = 0$$

Finally, substituting Eq. (A-15) into Eq. (A-14) gives the desired result, namely,

$$\sigma_\phi^2 \geq \frac{1}{K (2R_d)^2 (1 + 2R_d)} \quad (\text{A-16})$$

Note that at low SNR the bound approximately has an inverse square-law behavior with bit SNR as compared with the inverse linear behavior at high SNR.

It is important to emphasize that Eq. (A-16) is valid only when the denominator on the right-hand side of the equation is large. Thus, it is possible to apply the bound in Eq. (A-16) for small R_d provided that the number of bits in the observation, K , is sufficiently large.